Seeds | Backman
Stars and Galaxies
10th Edition

Universe Bowl

Imagine the history of the Universe as a time line down the middle of an American football field. The story begins on one goal line as the Big Bang fills the Universe with energy and a fantastically hot gas of hydrogen and helium. Follow the history from the first inch of the time line as the expansion of the Universe cools the gas and it begins to form galaxies and stars.

BIG BANG

The Dark Age when the Big Bang had cooled and before stars began to shine

Formation of the first galaxies well under way

The Age of Quasars: Galaxies, including our home galaxy, actively forming, colliding, and merging

The expansion of the Universe stops slowing and begins accelerating.

10 20 30 40 50

The First Inch

Goal line

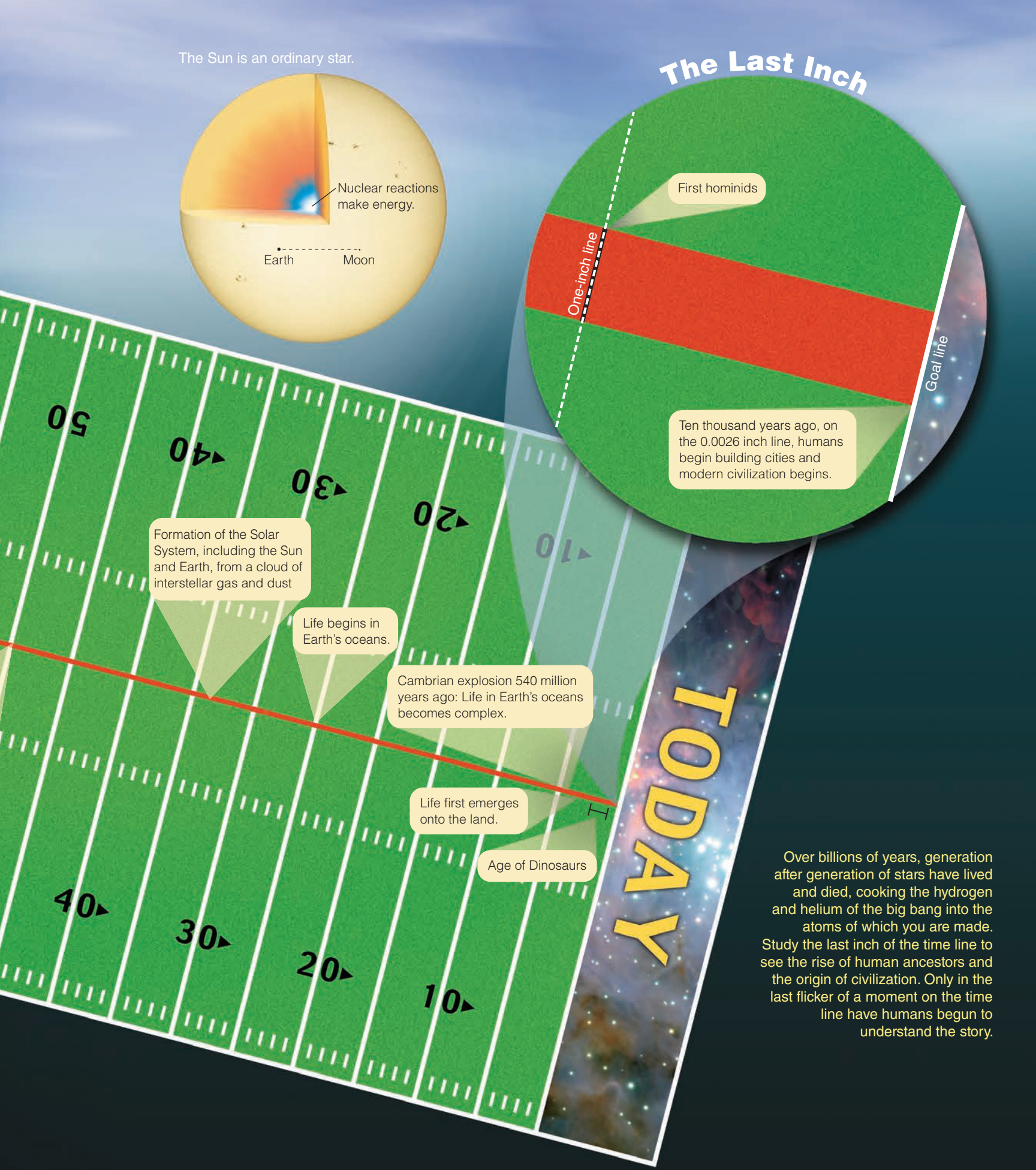
One-inch line

Recombination: A few hundred thousand years after the Big Bang, the gas becomes transparent to light.

Anglo-Australian Observatory/David Malin Images

A typical galaxy contains 100 billion stars.

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.


The Sun is an ordinary star.
Nuclear reactions make energy.
Earth
Moon
The Last Inch
First hominids
One-inch line
Goal line
Ten thousand years ago, on the 0.0026 inch line, humans begin building cities and modern civilization begins.
Formation of the Solar System, including the Sun and Earth, from a cloud of interstellar gas and dust
Life begins in Earth's oceans.
Cambrian explosion 540 million years ago: Life in Earth's oceans becomes complex.
Life first emerges onto the land.
Age of Dinosaurs
TODAY
50
40
30
20
10
Over billions of years, generation after generation of stars have lived and died, cooking the hydrogen and helium of the big bang into the atoms of which you are made. Study the last inch of the time line to see the rise of human ancestors and the origin of civilization. Only in the last flicker of a moment on the time line have humans begun to understand the story.

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Stars and Galaxies

TENTH EDITION

Michael A. Seeds

Joseph R. Grundy Observatory
Franklin and Marshall College

Dana E. Backman

SETI Institute & NASA Airborne Astronomy Ambassadors
SOFIA (Stratospheric Observatory for Infrared Astronomy)

CENGAGE

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit www.cengage.com/highered to search by ISBN#, author, title, or keyword for materials in your areas of interest.

Important Notice: Media content referenced within the product description or the product text may not be available in the eBook version.

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

***Stars and Galaxies,* Tenth Edition**
Michael A. Seeds, Dana E. Backman

Product Director: Dawn Giovanniello

Product Manager: Rebecca Berardy Schwartz

Content Developer: Rebecca Heider

Product Assistant: Caitlin Ghegan

Senior Marketing Manager: Heather R. Thompson

Senior Content Project Manager: Tanya Nigh

Senior Art Director: Cate Barr

Production Service: Edward J. Dionne, MPS Limited

Intellectual Property Analyst: Christine Myaskovsky

Intellectual Property Project Manager: Erika Mugavin

Photo Researcher: Geeta Kamath, Lumina Datamatics

Text Researcher: Karthick Govindaraju

Text and Cover Designer: tani hasegawa, tt eye

Cover and Title Page Image: Mission: *Hubble Space Telescope (HST)*; target: *Stephan's Quintet or galaxies NGC 7319*, *NGC 7318B and NGC 7318A*; spacecraft: *Hubble Space Telescope*; instrument: *Wide Field Planetary Camera 2*.

NASA, Jayanne English (University of Manitoba), Sally Hunsberger (Pennsylvania State University), Zolt Levay (Space Telescope Science Institute), Sarah Gallagher (Pennsylvania State University), and Jane Charlton

Compositor: MPS Limited

© 2019, 2016, 2013 Cengage Learning, Inc.

WCN: 02-300

Unless otherwise noted, all content is © Cengage.

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced or distributed in any form or by any means, except as permitted by U.S. copyright law, without the prior written permission of the copyright owner.

For product information and technology assistance, contact us at
Cengage Customer & Sales Support, 1-800-354-9706.

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2017945322

Softbound Edition:

ISBN: 978-1-337-39994-4

Loose-leaf Edition:

ISBN: 978-1-337-40010-7

Cengage
20 Channel Center Street
Boston, MA 02210
USA

Cengage is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com**.

Cengage products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage platforms and services, visit **www.cengage.com**.
To register or access your online learning solution or purchase materials for your course, visit **www.cengagebrain.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2017

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Dedication

In memory of Edward & Antonette Backman and Emery & Helen Seeds

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

BRIEF CONTENTS

PART 1: EXPLORING THE SKY

1 Here and Now 1

2 A User's Guide to the Sky 11

3 Moon Phases and Eclipses 32

4 Origins of Modern Astronomy 52

5 Gravity 78

6 Light and Telescopes 101

PART 2: THE STARS

7 Atoms and Spectra 127

8 The Sun 143

9 The Family of Stars 170

10 The Interstellar Medium 199

11 The Formation and Structure of Stars 216

12 Stellar Evolution 238

13 The Deaths of Stars 262

14 Neutron Stars and Black Holes 286

PART 3: THE UNIVERSE

15 The Milky Way Galaxy 310

16 Galaxies 335

17 Active Galaxies and Supermassive Black Holes 361

18 Modern Cosmology 377

PART 4: LIFE

19 Astrobiology: Life on Other Worlds 408

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CONTENTS

PART 1: EXPLORING THE SKY

1 Here and Now 1

1-1 Where Are We? 2
1-2 When Is Now? 6
1-3 Why Study Astronomy? 7

2 A User's Guide to the Sky 11

2-1 Stars and Constellations 12
2-2 The Sky and Celestial Motions 17
2-3 Sun and Planets 21
2-4 Astronomical Influences on Earth's Climate 25

3 Moon Phases and Eclipses 32

3-1 The Changeable Moon 33
3-2 Lunar Eclipses 36
3-3 Solar Eclipses 38
3-4 Predicting Eclipses 45

4 Origins of Modern Astronomy 52

4-1 Roots of Astronomy 53
4-2 The Copernican Revolution 58
4-3 Tycho, Kepler, and Planetary Motion 64
4-4 Galileo's Conclusive Evidence 70
4-5 Ninety-Nine Years That Revolutionized Astronomy 74

5 Gravity 78

5-1 Galileo's and Newton's Two New Sciences 79
5-2 Orbital Motion and Tides 84
5-3 Einstein and Relativity 92

6 Light and Telescopes 101

6-1 Radiation: Information from Space 102
6-2 Telescopes 105
6-3 Observatories on Earth: Optical and Radio 110
6-4 Airborne and Space Observatories 116
6-5 Astronomical Instruments and Techniques 118
6-6 Non-Electromagnetic Astronomy 123

HOW DO WE KNOW?

1-1 The Scientific Method 8
2-1 Scientific Models 17
2-2 Pseudoscience 26
2-3 Scientific Arguments 28
3-1 Scientific Imagination 39
4-1 Scientific Revolutions 64
4-2 Hypotheses, Theories, and Laws 69
5-1 Cause and Effect 83
5-2 Testing a Hypothesis by Prediction 92
6-1 Resolution and Precision 109

CONCEPT ART

The Sky Around You 18–19
The Cycle of the Seasons 22–23
The Phases of the Moon 34–35
An Ancient Model of the Universe 60–61
Orbits 86–87
Modern Optical Telescopes 112–113

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

PART 2: THE STARS

7 Atoms and Spectra 127

7-1 Atoms 128
7-2 Interactions of Light and Matter 131
7-3 Understanding Spectra 135

8 The Sun 143

8-1 The Solar Photosphere and Atmosphere 144
8-2 Solar Activity 151
8-3 Nuclear Fusion in the Sun 161

9 The Family of Stars 170

9-1 Star Distances 171
9-2 Apparent Brightness, Intrinsic Brightness, and Luminosity 173
9-3 Stellar Spectra 175
9-4 Star Sizes 179
9-5 Star Masses—Binary Stars 185
9-6 A Census of the Stars 191

10 The Interstellar Medium 199

10-1 Studying the Interstellar Medium 200
10-2 Components of the Interstellar Medium 208
10-3 The Gas-Stars-Gas Cycle 212

11 The Formation and Structure of Stars 216

11-1 Making Stars from the Interstellar Medium 217
11-2 The Orion Nebula: Evidence of Star Formation 221
11-3 Young Stellar Objects and Protostellar Disks 226
11-4 Stellar Structure 230
11-5 The Source of Stellar Energy 232

12 Stellar Evolution 238

12-1 Main-Sequence Stars 239
12-2 Post-Main-Sequence Evolution 246
12-3 Star Clusters: Evidence of Stellar Evolution 252
12-4 Variable Stars: Evidence of Stellar Evolution 253

13 The Deaths of Stars 262

13-1 Low-Mass Stars 263
13-2 The Evolution of Binary Stars 270
13-3 High-Mass Stars 273
13-4 Supernova Explosions 274
13-5 The End of Earth 282

14 Neutron Stars and Black Holes 286

14-1 Neutron Stars 287
14-2 Black Holes 297
14-3 Compact Objects with Disks and Jets 303

Celestial Profile **THE SUN** 144

HOW DO WE KNOW?

7-1 Quantum Mechanics 130
8-1 Confirmation and Consolidation 157
8-2 Scientific Confidence 165
9-1 Chains of Inference 186
9-2 Basic Scientific Data 194
10-1 Separating Facts from Hypotheses 209
11-1 Theories and Proof 221
12-1 Mathematical Models 241
13-1 Toward Ultimate Causes 268
14-1 Checks on Fraud in Science 301

CONCEPT ART

Atomic Spectra 136–137
Sunspots and the Solar Magnetic Cycle 154–155
Solar Activity and the Sun–Earth Connection 158–159
The Family of Stars 192–193
Three Kinds of Nebulae 202–203
Star Formation in the Orion Nebula 224–225
Observations of Young Stellar Objects and Protostellar Disks 228–229
Star Clusters and Stellar Evolution 254–255
Formation of Planetary Nebulae and White Dwarfs 266–267
The Lighthouse Model of Pulsars 290–291

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

PART 3: THE UNIVERSE

15 The Milky Way Galaxy 310

15-1 Discovery of the Galaxy 311
15-2 Structure of the Galaxy 315
15-3 Spiral Arms and Star Formation 319
15-4 The Nucleus of the Galaxy 323
15-5 Origin and Evolution of the Galaxy 325

16 Galaxies 335

16-1 The Family of Galaxies 336
16-2 Measuring the Properties of Galaxies 339
16-3 Evolution of Galaxies 349

17 Active Galaxies and Supermassive Black Holes 361

17-1 Active Galactic Nuclei 362
17-2 Supermassive Black Holes 369
17-3 A History of Galaxies and Supermassive Black Holes 372

18 Modern Cosmology 377

18-1 Introduction to the Universe 378
18-2 The Big Bang Theory 381
18-3 Space and Time, Matter and Gravity 389
18-4 21st-Century Cosmology 397

PART 4: LIFE

19 Astrobiology: Life on Other Worlds 408

19-1 The Nature of Life 409
19-2 Life in the Universe 411
19-3 Intelligent Life in the Universe 420

Afterword A-1
Appendix A Units and Astronomical Data A-3
Appendix B Observing the Sky A-11
Answers to Even-Numbered Problems AN-1
Glossary G-1
Index I-1

HOW DO WE KNOW?

15-1 Calibration 314
15-2 Nature as Processes 323
16-1 Classification in Science 339
17-1 Statistical Evidence 364
18-1 Reasoning by Analogy 380
18-2 Science: A System of Knowledge 388
18-3 Wishing Doesn't Make It So 403
19-1 UFOs and Space Aliens 421
19-2 The Copernican Principle 424

CONCEPT ART

Sagittarius A* 326–327
Galaxy Classification 340–341
Interacting Galaxies 352–353
Cosmic Jets and Radio Lobes 366–367
The Nature of Space-Time 392–393
DNA: The Code of Life 412–413

From Dana and Mike

We are excited that you are taking an astronomy course and using our book and related online resources. You are going to see and learn about some amazing things, from the icy rings of Saturn to monster black holes. We are proud to be your guides as you explore.

We have developed this book to help you expand your knowledge of astronomy, from recognizing the Moon and a few stars in the evening sky, to a deeper understanding of the extent, power, and diversity of the Universe. You will meet worlds where it rains methane, stars so dense their atoms are crushed, colliding galaxies that are ripping each other apart, and a Universe that is expanding faster and faster.

Two Goals

This book is designed to help you answer two important questions:

- **What are we?**
- **How do we know?**

By the question *"What are we?"* we mean: How do we fit into the Universe and its history? The atoms you are made of were born in the Big Bang when the Universe began, but those atoms were cooked and remade inside stars, and now they are inside you. Where will they be in a billion years? Astronomy is the only course on campus that can tell you that story, and it is a story that everyone should know.

By the question *"How do we know?"* we mean: How does science work? What is the evidence, and how do we use it? For instance, how can anyone know there was a Big Bang? In today's world, you need think carefully about the quality of the information with which we are flooded. You should demand evidence, not just explanations. Scientists have a special way of knowing based on evidence that makes scientific knowledge much more powerful than just opinion, policy, marketing, or public relations. It is the human race's best understanding of nature. To comprehend the world around you, you need to understand how science works. Throughout this book, you will find boxes called *How Do We Know?* and *Practicing Science.* They will help you understand how scientists use the methods of science to know what the Universe is like.

Expect to Be Astonished

One reason astronomy is exciting is that astronomers discover new things every day. Astronomers expect to be astonished. You can share in the excitement because we have worked hard to include new images, new discoveries, and new insights that will take you, in an introductory course, to the frontier of human knowledge. Telescopes on remote mountaintops and in space provide a daily dose of excitement that goes far beyond entertainment. These new discoveries in astronomy are exciting because they are about us. They tell us more and more about what we are.

As you read this book, notice that it is not organized as lists of facts for you to memorize. Rather, this book is organized to show you how scientists use evidence and theory to create logical arguments that explain how nature works. Look at the list of special features that follows this note. Those features were carefully designed to help you understand astronomy as evidence and theory. Once you see science as logical arguments, you hold the key to the Universe.

Don't Be Humble

As teachers, our quest is simple. We want you to understand your place in the Universe—your location not just in space but in the unfolding history of the physical Universe. We want you not only to know where you are and what you are in the Universe but also to understand how scientists know. By the end of this book, we want you to know that the Universe is very big but that it is described and governed by a small set of rules and that we humans have found a way to figure out the rules—a method called science.

To appreciate your role in this beautiful Universe, you need to learn more than just the facts of astronomy. You have to understand what we are and how we know. Every page of this book reflects that ideal.

Dana Backman
dbackman@seti.org

Mike Seeds
mseeds@fandm.edu

Key Content and Pedagogical Changes for the Tenth Edition

- Every chapter has been revised and updated with new text and images regarding observatories, scientific missions, and new discoveries.
- Some sections have been reorganized and updated to give clearer and more current presentations, especially Section 11-3, Young Stellar Objects and Protostellar Disks; and Section 13-4, Supernovae Explosions.
- Other chapters and sections with less substantial but still significant revisions are Section 6-3a, Modern Optical Telescopes; Section 6-6b, Gravity Wave Astronomy; Section 8-1d Composition of the Sun; and Section 9-1c.
- A *How Do We Know?* essay has been added to the cosmology chapter regarding what we can learn from a mistaken claim about discovering gravity waves,
- All art has been carefully reviewed and revised in the interest of accuracy, clarity, and consistency.
- *Key Equations* now have numbers and titles, as well as examples to demonstrate their use. This feature highlights and reinforces the equations that will be needed to solve problems in later chapters.
- New *Sense of Proportion* questions have been added to help students ground their understanding of relative sizes of celestial objects, distances, and so on. As a complement to this, discussion of proportionality has been expanded where relevant throughout the text.
- The *Focus on Fundamentals* boxed feature has been eliminated and relevant material was moved back into the text.
- The end-of-chapter *Summary* sections have been significantly revised in all chapters with the goal of helping students focus and better navigate key concepts for review.
- Discussion Questions have been replaced with *Active Inquiry Questions*, designed to engage students in deeper critical thinking.
- All numerical values in the text and tables were checked and in some cases updated.
- The features known as *Doing Science* in earlier editions were renamed *Practicing Science*.

Special Features

- The opening page of each chapter helps students see the organization of the book by connecting the chapter with the preceding and following chapters. The short list of important questions highlights the learning objectives of the chapter.
- *How Do We Know?* boxes help students understand how science works. Topics include the difference between a hypothesis and a theory, the use of statistical evidence, the construction of scientific models, and more.
- *Concept Art* features appear at least once in every chapter and cover topics that are strongly visual. Color and numerical keys in the introduction to the Concept Art guide you to the main concepts.
- *Common Misconceptions* highlight and correct popularly held misconceptions about astronomy.
- *Practicing Science* boxes at the end of most chapter sections begin with questions designed to put students into the role of scientists considering how best to proceed as they investigate the cosmos. These questions serve a second purpose as a further review of how we know what we know. Many of the *Practicing Science* boxes end with a second question that points the student-as-scientist in a direction for further investigation.
- *Celestial Profiles* directly compare and contrast planets with each other. This is the way planetary scientists understand the planets—not as isolated, unrelated bodies but as siblings; they have noticeable differences but many characteristics and a family history in common.
- *What Are We?* features at the end of each chapter show how the chapter content helps explain our place in the cosmos.
- Chapter *Summaries* review the key concepts of the chapter, highlighting key terms and equations, to aid student study.
- End-of-chapter *Review Questions* are designed to help students review and test their understanding of the material.
- End-of-chapter *Active Inquiry Questions* go beyond the text and invite students to think critically and creatively about scientific questions.
- End-of-chapter *Problems* promote quantitative understanding of the chapter contents.
- *Sense of Proportion* questions gauge student understanding of relative sizes and quantities in the Universe.
- *Learning to Look* questions prompt students to answer questions based on observations of visual evidence shown in diagrams or photographs.

MindTap for Astronomy

MindTap Astronomy for *Stars and Galaxies*, 10th Edition, is the digital learning solution that powers students from memorization to mastery. It gives you complete control of your course—to

provide engaging content, to challenge every individual and to build their confidence. Empower students to accelerate their progress with MindTap. MindTap: Powered by You.

MindTap for Astronomy has a carefully curated learning path that includes tutorial simulations, readings, and assessments. Research has proven that students perform better when activities encourage an active experience; with this research in mind, author Mike Seeds has developed tutorial simulations that are integrated right into the MindTap reader to help students better visualize the concepts. Animation tutorials will build student reasoning so they will ultimately be able to draw stronger conclusions.

The end-of-chapter homework questions in MindTap provide a tighter integration with the textbook content and emphasize conceptual understanding.

Virtual Astronomy Labs 3.0 used real astronomical data combined with robust simulations in auto-graded modular segments to provide a true online laboratory experience. Through the use of simulations and exercises, students are introduced to fundamental and complex theories. The labs focus on 20 of the most important concepts in astronomy, such as astronomical measurements, dark matter, black holes, binary stars, and extrasolar planets.

Waypoints bring the Concept Art from the printed page alive in short (4- to 6-minute) animated lessons narrated by author Mike Seeds. Waypoints serve to clarify important concepts and present dynamic interactions in full multimedia presentations on topics ranging from the Celestial Sphere and the Ancient Universe to Emission Nebulae, Galaxy Classification, and Stellar Evolution.

ACKNOWLEDGMENTS

Over the years, we have had the guidance of a great many people who care about astronomy and teaching. We would like to thank all of the students and teachers who have contributed to this book. They helped shape the book through their comments and suggestions.

Many observatories, research institutes, laboratories, and individual astronomers have supplied figures and diagrams for this edition. They are listed on the credits page, and we would like to thank them specifically for their generosity.

We are happy to acknowledge the use of images and data from a number of important programs. In preparing materials for this book we used several atlas images and mosaics produced by the Two Micron All Sky Survey (2MASS), a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/JPL-Caltech, funded by NASA and the NSF. A number of solar images are used courtesy of the *SOHO* consortium, a project of international cooperation between ESA and NASA. The NASA Sky View facility located at Goddard Space Flight Center and the SIMBAD database operated by the CDS in Strasbourg, France, were also used in preparation of this text.

It has been a great pleasure to work with our Cengage production team, Content Developer Rebecca Heider, Product Assistant Caitlin Ghegan, Art Director Cate Barr, and Product Manager Rebecca Berardy-Schwartz, plus Edward Dionne of MPS Limited.

Most of all, we would like to thank our families for putting up with "the books." They know all too well that textbooks are made of time.

Dana Backman
Mike Seeds

ABOUT THE AUTHORS

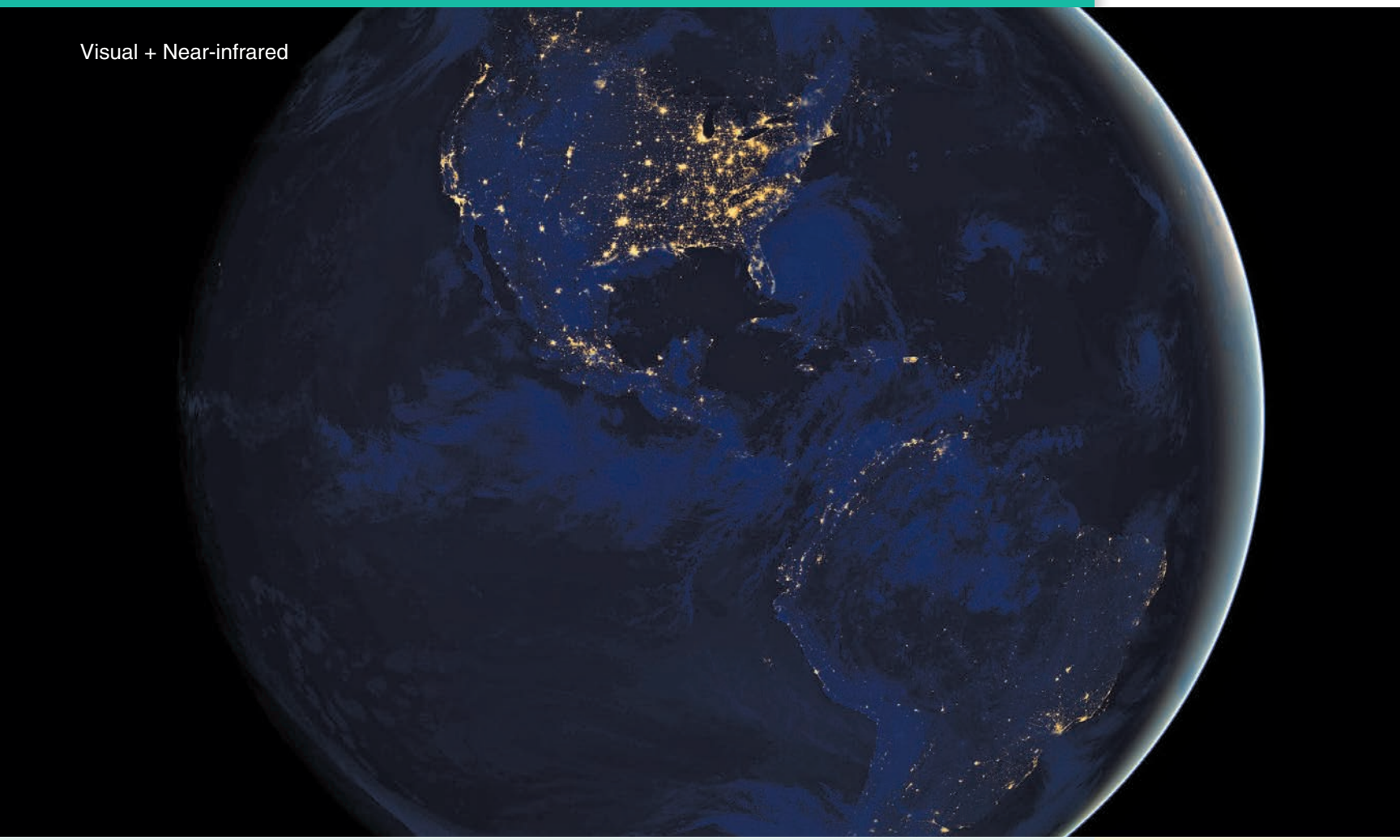
Seth Shostak / SETI Institute

Dana Backman taught in the physics and astronomy department at Franklin and Marshall College in Lancaster, Pennsylvania, from 1991 until 2003. He invented and taught a course titled "Life in the Universe" in F&M's interdisciplinary Foundations program. Dana also has taught introductory astronomy and astronomy for physics majors at Santa Clara University as well as introductory astronomy, astrobiology, and cosmology courses in Stanford University's Continuing Studies program. His research interests focused on infrared observations of planet formation, models of debris disks around nearby stars, and evolution of the Solar System's Kuiper Belt. Dana is employed by the SETI Institute in Mountain View, California, managing NASA's Airborne Astronomy Ambassadors and outreach programs for SOFIA, the Stratospheric Observatory for Infrared Astronomy. Along with this book, Dana is coauthor with Mike Seeds of *Horizons: Exploring the Universe; Foundations of Astronomy; The Solar System;* and *ASTRO*, all published by Cengage.

Courtesy of Kris Koenig

Mike Seeds was a professor of physics and astronomy at Franklin and Marshall College in Lancaster, Pennsylvania, from 1970 until his retirement in 2001. In 1989 he received F&M College's Lindback Award for Distinguished Teaching. Mike's love for the history of astronomy led him to create upper-level courses on archaeoastronomy and on the Copernican Revolution ("Changing Concepts of the Universe"). His research interests focused on variable stars and automation of astronomical telescopes. In addition to this book, Mike is coauthor with Dana Backman of *Horizons: Exploring the Universe; Foundations of Astronomy; The Solar System;* and *ASTRO*, all published by Cengage. He was senior consultant for creation of the 20-episode telecourse accompanying his book *Horizons: Exploring the Universe.*

1 Here and Now



Courtesy of NASA

▲ Composite image of Earth's western hemisphere at night assembled from data acquired by the *Suomi National Polar-orbiting Partnership* satellite. The bright crescent at right is the location of dawn.

As you study astronomy, you will learn about yourself. You are a planet-walker, and this chapter will give you a preview of what that means. The planet you live on whirls around a star that moves through a Universe filled with other stars and galaxies which are all results of billions of years of history and evolution. You owe it to yourself to know where you are in the Universe, and when you are in its history, because those are important steps toward knowing what you are.

In this chapter, you will consider three important questions about astronomy:

- Where is Earth in the Universe?
- How does human history fit into the history of the Universe?
- Why study astronomy?

This chapter is a jumping-off point for your exploration of deep space and deep time. The next chapter continues your journey by looking at the night sky as seen from Earth. As you study astronomy, you will see how science gives you a way to know how nature works. Later chapters will provide more specific insights into how scientists study and understand nature.

The longest journey begins with a single step. —LAOZI

1-1 Where Are We?

To find your place among the stars, you can take a cosmic zoom—a ride out through the Universe to preview the kinds of objects you are about to study.

Begin with something familiar. **Figure 1-1** shows an area 52 feet across on a college campus including a person, a sidewalk, and a few trees, which are all objects with sizes you can understand. Each successive picture in this "zoom" will show you a region of the Universe that is 100 times wider than the preceding picture. That is, each step will widen your **field of view**, which is the region you can see in the image, by a factor of 100.

Widening your field of view by a factor of 100 allows you to see an area 1 mile in diameter in the next image (**Figure 1-2**). People, trees, and sidewalks have become too small to discern, but now you can view an entire college campus plus surrounding streets and houses. The dimensions of houses and streets are familiar; this is still the world you know.

Before leaving this familiar territory, you need to change the units you use to measure sizes. All scientists, including astronomers, use the metric system of units because it is well understood worldwide and, more important, because it simplifies calculations. If you are not already familiar with the metric system, or if you need a review, study Appendix A before reading on.

In metric units, the image in Figure 1-1 is about 16 meters across, and the 1-mile diameter of Figure 1-2 equals about 1.6 kilometers. You can see that a kilometer (abbreviated km) is a bit less than two-thirds of a mile—a short walk across a neighborhood. When you expand your field of view by another factor of 100, the neighborhood you saw in Figure 1-2 vanishes. Now your field of view is 160 km wide, and you see cities and towns as patches of gray (**Figure 1-3**). Wilmington, Delaware, is visible at the lower right. At this scale, you can see some of the natural features of Earth's surface. The Allegheny Mountains of southern Pennsylvania cross the image at the upper left, and the Susquehanna River flows southeast into Chesapeake Bay. What look like white bumps are a few puffs of cloud.

Figure 1-3 is an infrared photograph in which healthy green leaves and crops are shown as red. Human eyes are sensitive to only a narrow range of colors. As you explore the Universe, you will learn to use a wide range of other "colors," from X-rays to radio waves, to reveal sights invisible to unaided human eyes. You will learn more about infrared, X-ray, and radio energy in later chapters.

At the next step in your journey, you can see your entire planet, which is nearly 13,000 km in diameter (**Figure 1-4**). At any particular moment, half of Earth's surface is exposed to sunlight, and the other half is in darkness. As Earth rotates on its axis, it carries you through sunlight and then through darkness, producing the cycle of day and night. The blurriness at the right edge of the Earth image is the boundary between day and night—the sunset line. This is a good example of how a photo can give you visual clues to understanding a concept. Special questions called "Learning to Look" at the end of each chapter give you a chance to use your own imagination to connect images with explanations about astronomical objects.

Enlarge your field of view by another factor of 100, and you see a region 1,600,000 km wide (**Figure 1-5**). Earth is the small blue dot in the center, and the Moon, the diameter of which is only one-fourth of Earth's, is an even smaller dot along its orbit

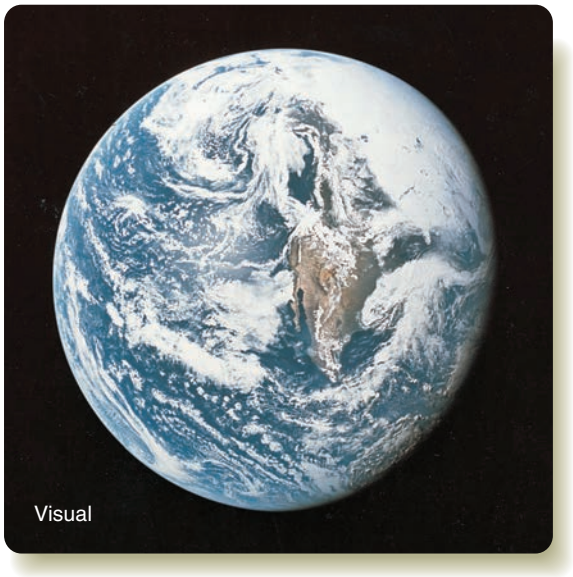
Michael A. Seeds

▲ **Figure 1-1** This familiar scene is an area about 52 feet (16 meters) wide.

Imagery ©2016 DigitalGlove, PA Department of Conservation and Natural Resources-PAMP /USGS, U.S. Geological Survey, USDA 500 ft Farm Service Agency, Map data ©2016 Google

▲ **Figure 1-2** The field of view is 1 mile (1.6 km) wide. This box ■ represents the relative size of Figure 1-1.

▲ **Figure 1-3** The field of view is 160 km wide. This box ■ represents the relative size of Figure 1-2.



▲ **Figure 1-4** The field of view is 16,000 km wide. This box ■ represents the relative size of Figure 1-3.

380,000 km away. (The relative sizes of Earth and Moon are shown in the inset at the bottom right of Figure 1-5.)

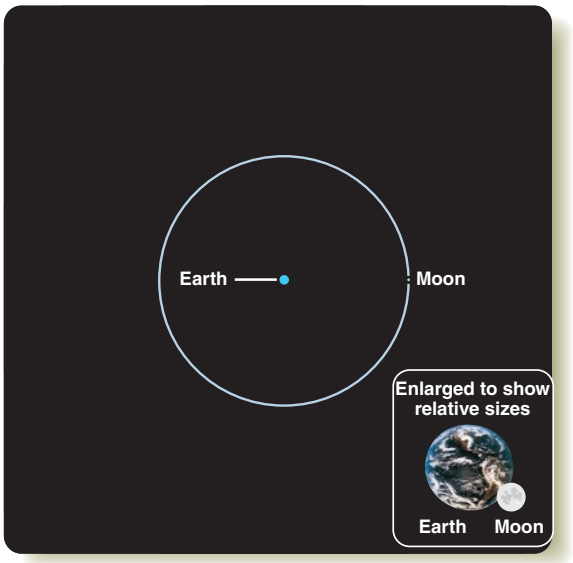
The numbers in the preceding paragraph are so large that it is inconvenient to write them out. Soon you will be using numbers even larger than these to describe the Universe; rather than writing such astronomical numbers as they are in the previous paragraph, it is more convenient to write them in **scientific notation**. This is nothing more than a simple way to write very big or very small numbers without using lots of zeros. For example, in scientific notation 380,000 becomes 3.8×10^5. If you are not familiar with scientific notation, read the section on "Powers of 10 Notation" in Appendix A. The Universe is too big to describe without using scientific notation.

When you once again enlarge your field of view by a factor of 100 (**Figure 1-6**), Earth, the Moon, and the Moon's orbit that filled the previous figure are indistinguishable in the blue dot at lower left of the new figure. Now you can see the Sun and two other planets that are part of our Solar System. Our **Solar System** consists of the Sun, its family of planets, and some smaller bodies such as moons, asteroids, and comets.

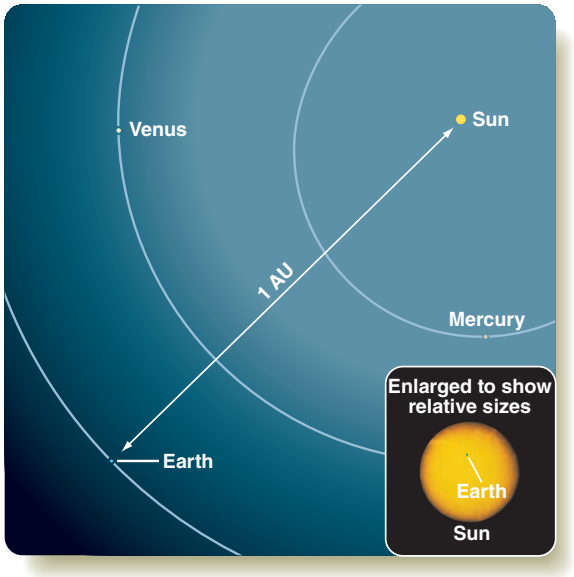
Earth, Venus, and Mercury are **planets**, which are spherical, nonluminous bodies that orbit a star and shine by reflected light. Venus is about the size of Earth, and Mercury has slightly more than one-third of Earth's diameter. On this diagram, they are both too small to be portrayed as anything but tiny dots. The Sun is a **star**, a self-luminous ball of hot gas. Even though the Sun is about 100 times larger in diameter than Earth (inset at bottom right of Figure 1-6), it, too, is no more than a dot in this diagram. Figure 1-6 represents an area with a diameter of 1.6×10^8 km.

Another way astronomers simplify descriptions and calculations that require large numbers is to define larger units of measurement. For example, the average distance from Earth to the Sun is a unit of distance called the **astronomical unit (AU)**; an AU is equal to 1.5×10^8 km. Using that term, you can express the average distance from Mercury to the Sun as about 0.39 AU and the average distance from Venus to the Sun as about 0.72 AU.

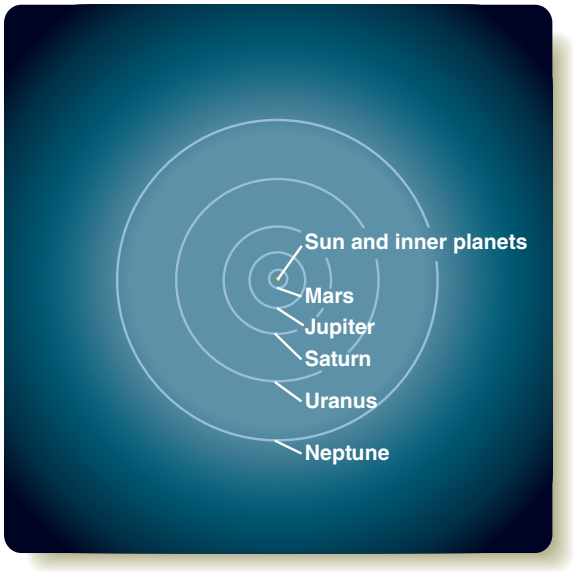
These distances are averages because the orbits of the planets are not perfect circles. This is especially apparent in the case of Mercury. Its orbit carries it as close to the Sun as 0.31 AU and as far away as 0.47 AU. You can see the variation in the distance from Mercury to the Sun in Figure 1-6. Earth's orbit is more circular than Mercury's; its distance from the Sun varies by only a few percent.



▲ **Figure 1-5** The field of view is 1.6 million km wide. This box ■ represents the relative size of Figure 1-4.



▲**Figure 1-6** The field of view is 160 million km wide. This box ■ represents the relative size of Figure 1-5.



▲**Figure 1-7** The field of view is 16 billion km (about 110 AU) wide. The orbits of the three innermost planets—Mercury, Venus, and Earth—are too small to show at this scale. This box ■ represents the relative size of Figure 1-6.

Enlarge your field of view again by a factor of 100, and you can see the entire planetary region of our Solar System (**Figure 1-7**). The Sun, Mercury, Venus, and Earth lie so close together that you cannot see them separately at this scale. You can see only the brighter, more widely separated objects such as Mars, the next planet outward. Mars is only 1.5 AU from the Sun, but Jupiter, Saturn, Uranus, and Neptune are farther from the Sun, so they are easier to locate in this diagram. They are cold worlds that are far from the Sun's warmth. Light from the Sun reaches Earth in only 8 minutes, but it takes more than 4 hours to reach Neptune.

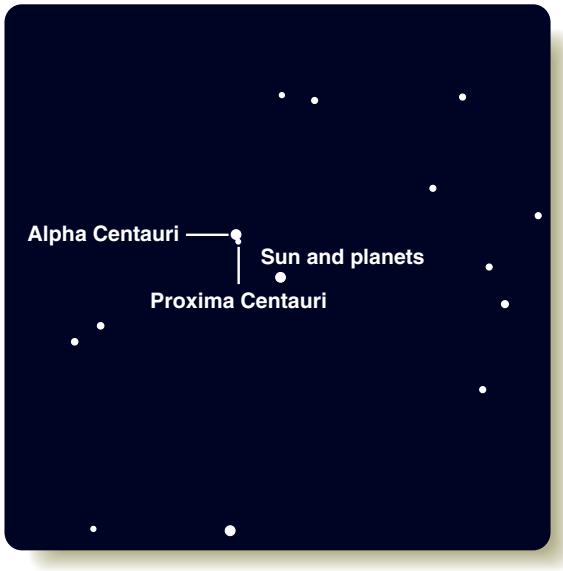
You can remember the order of the planets from the Sun outward by remembering a simple sentence such as: *My Very Educated Mother Just Served Us Noodles* (perhaps you can come up with a better one). The first letter of each word is the same as the first letter of a planet's name: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. The list of planets once included Pluto, but in 2006, astronomers attending an international scientific congress made the decision that Pluto should be redefined as a **dwarf planet**. Pluto meets some of the criteria to be considered a planet, but not others. Pluto is now known to be just one of a group of small objects that have been discovered circling the Sun beyond Neptune.

When you again enlarge your field of view by a factor of 100, the Solar System vanishes (**Figure 1-8**). The Sun is only a point of light, and all the planets and their orbits are now crowded into the small yellow dot at the center. The planets are too small and too faint to be visible so near the brilliance of the Sun.

Notice that no stars are visible in Figure 1-8 except for the Sun. The Sun is a fairly typical star, and it seems to be located in a fairly average neighborhood in the Universe. Although there are many billions of stars like the Sun, none is close enough to be visible in this diagram, which shows a region only 11,000 AU in diameter. Stars in the Sun's neighborhood are typically separated by distances about 30 times larger than that.

In **Figure 1-9**, your field of view has expanded again by a factor of 100 to a diameter of 1.1 million AU. The Sun is at the center, and at this scale you can see a few of the nearest stars. These stars are so distant that it is not convenient to give their distances in AU. To express distances so large, astronomers defined a new unit of distance, the light-year. One **light-year (ly)** is the distance that light travels in one year, approximately 9.5×10^{12} km or 63,000 AU. It is a **Common Misconception** that a light-year is a unit of time,

▲**Figure 1-8** The field of view is 11,000 AU wide. This box ■ represents the relative size of Figure 1-7.



▲ **Figure 1-9** The field of view is 17 ly wide. The sizes of the dots representing each star are not to scale. This box ■ represents the relative size of Figure 1-8.

NSF/AURA/NOAO/Bill Schoening

▲ **Figure 1-10** The field of view is 1700 ly wide. This box ■ represents the relative size of Figure 1-9.

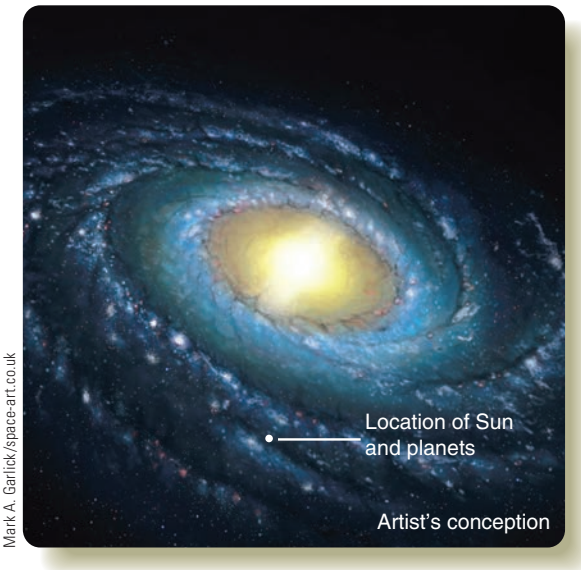
and you can sometimes hear the term misused in science fiction movies and TV shows. The next time you hear someone say, "It will take me light-years to finish my history paper," you could tell the person that a light-year is a distance, not a time (although perhaps that comment wouldn't be appreciated). The diameter of your field of view in Figure 1-9 is 17 ly.

Another **Common Misconception** is that stars look like disks when seen through a telescope. Although most stars are approximately the same size as the Sun, they are so far away that astronomers cannot see them as anything but points of light. Even the closest star to the Sun—Proxima Centauri, which is only 4.2 ly from Earth—looks like a point of light through the biggest telescopes on Earth. Figure 1-9 follows the common astronomical practice of making the sizes of the dots represent not the sizes of the stars but their brightness. This is how star images are recorded on photographs. Bright stars make larger spots on a photograph than faint stars, so the size of a star image in a photo tells you not how big the star is but rather how bright it is.

You might wonder whether other stars have families of planets orbiting around them as the Sun does. Such objects, termed *extrasolar planets,* are very difficult to see because they are generally small, faint, and too close to the glare of their respective parent stars. Nevertheless, astronomers have used indirect methods to find thousands of such objects, although only a handful have been photographed directly.

In **Figure 1-10**, you expand your field of view by another factor of 100, and the Sun and its neighboring stars vanish into the background of thousands of other stars. The field of view is now 1700 ly in diameter. Of course, no one has ever journeyed thousands of light-years from Earth to look back and photograph our neighborhood, so this is a representative photograph of the sky. The Sun is a relatively faint star that would not be easily located in a photo at this scale.

If you again expand your field of view by a factor of 100, you see our Galaxy, with a visible disk of stars about 80,000 ly in diameter (**Figure 1-11**). A **galaxy** is a great cloud of stars, gas, and dust held together by the combined gravity of all of its matter. Galaxies range from 1000 ly to more than 300,000 ly in diameter, and the biggest ones contain more than a trillion (10^{12}) stars. In the night sky, you can see our Galaxy as a great, cloudy wheel of stars ringing the sky. This band of stars is known as the *Milky Way,* and our home galaxy is called the **Milky Way Galaxy**.



Mark A. Garlick/space-art.co.uk

▲ **Figure 1-11** The field of view is 170,000 ly wide. This box ■ represents the relative size of Figure 1-10.

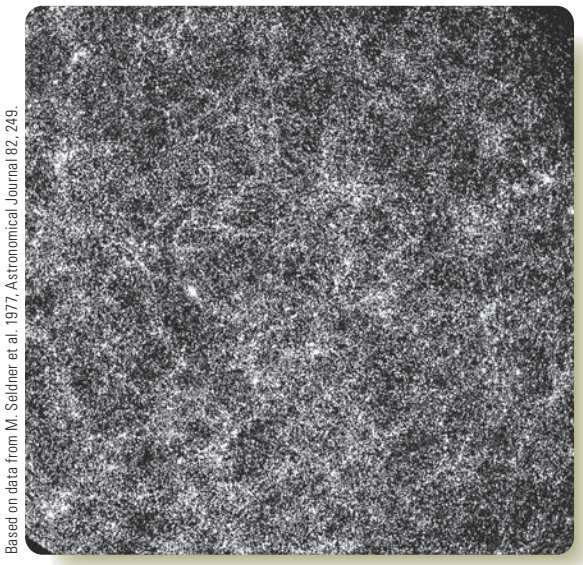
How does anyone know what the disk of the Milky Way Galaxy would look like from a vantage point tens of thousands of light years away? Astronomers use evidence to guide their explanations as they envision what our Galaxy looks like. Artists can then use those scientific descriptions to create a painting. Many images in this book are artists' conceptions of objects and events that are too big or too dim to see clearly, emit energy your eyes cannot detect, or happen too slowly or too rapidly for humans to sense. These images are much better than guesses; they are illustrations guided by the best scientific information astronomers can gather. As you continue to explore, notice how astronomers use the methods of science to understand and depict cosmic events.

The artist's conception of the Milky Way Galaxy reproduced in Figure 1-11 shows that our Galaxy, like many others, has graceful spiral arms winding outward through its disk. In a later chapter, you will learn that the spiral arms are places where stars are formed from clouds of gas and dust. Our own Sun was born in one of these spiral arms, and, if you could see the Sun in this picture, it would be in the disk of the Galaxy about two-thirds of the way out from the center, at about the location of the marker dot indicated in the figure.

Ours is a fairly large galaxy. Only a century ago astronomers thought it was the entire Universe—an island cloud of stars in an otherwise empty vastness. Now they know that the Milky Way Galaxy is not unique; it is only one of many billions of galaxies scattered throughout the Universe.

You can see a few of these other galaxies when you expand your field of view by another factor of 100 (**Figure 1-12**). Our Galaxy appears as a tiny luminous speck surrounded by other specks in a region 17 million light-years in diameter. Each speck represents a galaxy. Notice that our Galaxy is part of a group of a few dozen galaxies. Galaxies are commonly grouped together in such clusters. Some galaxies have beautiful spiral patterns like our home, the Milky Way Galaxy, some are globes of stars without spirals, and some seem strangely distorted. In a later chapter, you will learn what produces these differences among the galaxies.

▲ **Figure 1-12** The field of view is 17 million ly wide. This box ■ represents the relative size of Figure 1-11.



Based on data from M. Seldner et al. 1977, Astronomical Journal 82, 249.

▲ **Figure 1-13** The field of view is 1.7 billion ly wide. This box ■ represents the relative size of Figure 1-12.

Now is a chance for you to spot another **Common Misconception**. People often say *Galaxy* when they mean *Solar System*, and they sometimes confuse both terms with *Universe*. Your cosmic zoom has shown you the difference. The Solar System is your local neighborhood, that is, the Sun and its planets, one planetary system. The Milky Way Galaxy contains our Solar System plus billions of other stars and whatever planets orbit around them—in other words, billions of planetary systems. The Universe includes everything: all of the galaxies, stars, and planets, including our Galaxy and, a very small part of that, our Solar System. Distinguishing among the Solar System, the Galaxy, and the Universe requires having an accurate sense of proportion.

If you expand your field of view one more time, you can see that clusters of galaxies are connected in a vast network (**Figure 1-13**). Clusters are grouped into superclusters—clusters of clusters—and the superclusters are linked to form long filaments and walls outlining nearly empty voids. These filaments and walls appear to be the largest structures in the Universe. Were you to expand your field of view another time, you would probably see a uniform fog of filaments and walls. When you puzzle over the origin of these structures, you are at the frontier of human knowledge.

1-2 When Is Now?

Now that you have an idea where you are in space, you might also like to know where you are in time. The stars shone for billions of years before the first human looked up and wondered

what they were. To get a sense of your place in time, all you need is a long ribbon.

Imagine stretching that ribbon from goal line to goal line down the center of a U.S. football field, a distance of 100 yards (about 91 meters), as shown on the inside front cover of the printed book. Imagine that one end of the ribbon represents *today*, and the other end represents the beginning of the Universe—the moment that astronomers call the *Big Bang*. In Chapter 18, Modern Cosmology, you will learn about the Big Bang and evidence that the Universe is approximately 14 billion years old. Your ribbon represents 14 billion years, the entire history of the Universe.

Imagine beginning at the goal line labeled *BIG BANG* and replaying the entire history of the Universe as you walk along your ribbon toward the goal line labeled *TODAY*. Astronomers have evidence that the Big Bang initially filled the entire Universe with hot, glowing gas, but, as the gas cooled and dimmed, the Universe went dark. That all happened along the first half-inch of the ribbon. There was no light for the next 400 million years, until gravity was able to pull some of the gas together to form the first stars. That seems like a lot of years, but if you stick a little flag beside the ribbon to mark the birth of the first stars, it would be not quite 3 yards from the goal line where the Universe's history began.

You have to walk only about 4 or 5 yards along the ribbon before galaxies formed in large numbers. Our home galaxy would be one of those taking shape. By the time you cross the 50-yard line, the Universe is full of galaxies, but the Sun and Earth have not formed yet. You need to walk past the 50-yard line all the way to the other 33-yard line before you can finally stick a flag beside the ribbon to mark the formation of the Sun and planets—our Solar System—4.6 billion years ago and about 9 billion years after the Big Bang.

You can carry your flags a few yards further to about the 25-yard line, 3.4 billion years ago, to mark the earliest firm evidence for life on Earth—microscopic creatures in the oceans—and you have to walk all the way to the 3-yard line before you can mark the emergence of life on land only 0.4 billion (400 million) years ago. Your dinosaur flag goes inside the 2-yard line. Dinosaurs go extinct as you pass the one-half-yard line, 65 million years ago.

What about people? You can put a little flag for the first humanlike creatures, 4 million years ago, only about 1 inch (2.5 cm) from the goal line labeled *TODAY*. Civilization, the building of cities, began about 10,000 years ago, so you have to try to fit that flag in only 0.0026 inch from the goal line. That's less than the thickness of the page you are reading right now. Compare the history of human civilization with the history of the Universe. Every war you have ever heard of, the life of every person whose name is recorded, and the construction of every structure ever made, from Stonehenge to the building you are in right now, fits into that 0.0026 inch of the time ribbon.

Humanity is very new to the Universe. Our civilization on Earth has existed for only a flicker of an eyeblink in the history of the Universe. As you will discover in the chapters that follow, only in the last hundred years or so have astronomers begun to understand where we are in space and in time.

1-3 Why Study Astronomy?

Your exploration of the Universe will help you answer two fundamental questions:

What are we?
How do we know?

The question "What are we?" is the first organizing theme of this book. Astronomy is important to you because it will tell you what you are. Notice that the question is not "*Who* are we?" If you want to know who we are, you may want to talk to a sociologist, theologian, artist, or poet. "*What* are we?" is a fundamentally different question.

As you study astronomy, you will learn how you fit into the history of the Universe. You will learn that the atoms in your body had their birth in the Big Bang when the Universe began. Those atoms have been cooked and remade inside generations of stars, and now, after more than 10 billion years, they are inside you. Where will they be in another 10 billion years? This is a story everyone should know, and astronomy is the only course on campus that can tell you that story.

Every chapter in this book ends with a short feature titled "What Are We?" This summary shows how the astronomy in the chapter relates to your part in the story of the Universe.

The question "How do we know?" is the second organizing theme of this book. It is a question you should ask yourself whenever you encounter statements made by so-called experts in any field. Should you swallow a diet supplement recommended by a TV star? Should you vote for a candidate who denies that we face a climate crisis? To understand the world around you and to make wise decisions for yourself, for your family, and for your nation, you need to understand how science works.

You can use astronomy as a case study in science. In every chapter of this book, you will find short essays titled "How Do We Know?" They are designed to help you think not about *what* is known but about *how* it is known. To do that, these essays will explain different aspects of scientific thought processes and procedures to help you understand how scientists learn about the natural world.

Over the last four centuries, a way to understand nature has been developed that is called the **scientific method** (**How Do We Know? 1-1**). You will see this process applied over and over as you read about exploding stars, colliding galaxies, and alien planets. The Universe is very big, but it is described by a small set of rules, and we humans have found a way to figure out the rules by using a method called science.

HOW DO WE KNOW? 1-1

The Scientific Method

How do scientists learn about nature?

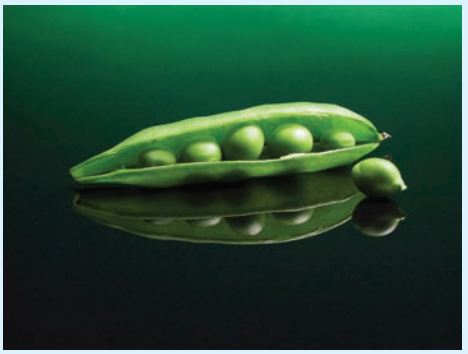
You have probably heard several times during your education about the scientific method as the process by which scientists form hypotheses and test them against evidence gathered by experiments and observations. That is an oversimplification of the subtle and complex ways that scientists actually work.

Scientists use the scientific method all the time, and it is critically important, but they rarely think of it while they are doing it, any more than you think about the details of what you are doing while you are riding a bicycle. It is such an ingrained way of thinking about and understanding nature that it is almost transparent to the people who use it most.

Scientists try to form hypotheses that explain how nature works. If a hypothesis is contradicted by evidence from experiments or observations, it must be revised or discarded. If a hypothesis is confirmed, it still must be tested further. In that very general way, the scientific method is a way of testing and refining ideas to better describe how nature works.

For example, Gregor Mendel (1822–1884) was an Austrian abbot who liked plants. He formed a hypothesis that offspring usually inherit traits from their parents not as a smooth blend, as most scientists of the time believed, but in discrete units according to strict mathematical rules. Mendel cultivated and tested more than 28,000 pea plants, noting which produced smooth peas and which produced wrinkled peas and how that trait was inherited by successive generations. His study of pea plants confirmed his hypothesis and allowed the development of a series of laws of inheritance. Although the importance of his work was not recognized in his lifetime, Mendel is now called the "father of modern genetics."

The scientific method is not a simple, mechanical way of grinding facts into understanding; a scientist needs insight and ingenuity both to form and to test good hypotheses. Scientists use the scientific method almost automatically, sometimes forming, testing, revising, and discarding hypotheses minute by minute as they discuss a new idea, other times spending years studying a single promising hypothesis.



Inspirestock/Jupiter Images

Using the scientific method, Gregor Mendel discovered that whether peas are smooth or wrinkled is an inherited trait.

The scientific method is, in fact, a combination of many ways of analyzing information, finding relationships, and creating new ideas, in order to know and understand nature. The "How Do We Know?" essays in the chapters that follow will introduce you to some of those techniques.

WHAT ARE WE?

Participants

Astronomy gives you perspective on what it means to be here on Earth. This chapter has helped you locate yourself in space and time. Once you realize how vast our Universe is, Earth seems quite small. People on the other side of the world seem like neighbors. And, in the entire history of the Universe, the story of humanity is only the blink of an eye. This may seem humbling at first, but you can be proud of how much we humans have understood in such a short time.

Not only does astronomy locate you in space and time, it places you within the physical processes that govern the Universe. Gravity and atoms work together to make stars, generate energy, light the Universe, and create the chemical elements in your body. The chapters that follow will show how you fit into that cosmic process.

Although you are very small and your kind have existed in the Universe for only a short time, you are an important participant in something very large and beautiful.

Study and Review

SUMMARY

1-1 Where Are We?

- You surveyed the Universe by taking a cosmic zoom in which each **field of view** was 100 times wider than the previous field of view.
- Astronomers use the metric system because it simplifies calculations, and they use **scientific notation** for very large or very small numbers.
- You live on a **planet**, Earth, which orbits our **star**, the Sun, once per year. As Earth rotates once per day, you see the Sun rise and set.
- The **Solar System** includes the Sun at the center, all of the major planets that orbit around it—Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune—plus the moons of the planets and other objects such as asteroids, comets, and **dwarf planets** like Pluto, bound to the Sun by its gravity.
- The **astronomical unit (AU)** is the average distance from Earth to the Sun. The **light-year (ly)** is the distance light can travel in one year.

- Astronomers have found thousands of planets orbiting stars other than our Sun, even though such distant and small bodies are very difficult to detect.
- The Milky Way, the hazy band of light that encircles the sky, is our **Milky Way Galaxy** seen from inside. **Galaxies** contain many billions of stars. The Milky Way Galaxy is about 80,000 ly in diameter and contains more than 100 billion stars including our Sun.
- Our Galaxy is just one of billions of galaxies that fill the Universe in great clusters, clouds, filaments, and walls—the largest structures in the Universe.

1-2 When Is Now?

- Astronomers have evidence that the Universe began about 14 billion years ago in an event called the Big Bang that filled the Universe with hot gas.
- The hot gas cooled, the first galaxies began to form, and stars began to shine about 400 million years after the Big Bang.
- The Sun and planets of our Solar System formed about 4.6 billion years ago.
- Life began in Earth's oceans soon after Earth formed but did not emerge onto land until 400 million years ago, less than 1/30 of the age of the Universe. Dinosaurs evolved relatively soon after that and went extinct just 65 million years ago.
- Humanlike creatures developed on Earth only about 4 million years ago, less than 1/3000 of the age of the Universe, and human civilizations developed just 10,000 years ago.

1-3 Why Study Astronomy?

- Although astronomy seems to be about stars and planets, it describes the Universe in which you live, so it is really about you. Astronomy helps you answer the question, "What are we?"
- As you study astronomy, you should ask, "How do we know?" and that will help you understand how science provides a way to understand nature.
- In its simplest outline, science follows the **scientific method**, in which scientists test hypotheses against evidence from experiments and observations. This method is a powerful way to learn about nature.

REVIEW QUESTIONS

1. The field of view in Figure 1-2 is a factor of 100 larger than the field of view in Figure 1-1. What aspects of Figure 1-2 increased by a factor of 100 relative to Figure 1-1? Did the height increase by that amount? The diameter? The area?
2. What is the largest dimension of which you have personal sensory experience? Have you ever hiked 10 miles? Run a marathon? Driven across a continent? Flown to the opposite side of Earth?
3. What is the difference between the Solar System, the Galaxy, and the Universe?
4. What is the difference between the Moon and a moon?
5. Why do astronomers now label Pluto a "dwarf planet"?
6. Why are light-years more convenient than miles, kilometers, or AU for measuring the distances to stars and galaxies?
7. Why is it difficult to detect planets orbiting other stars?
8. What does the size of the star image in a photograph tell you?
9. What is the difference between the Milky Way and the Milky Way Galaxy?
10. What are the Milky Way Galaxy's spiral arms?
11. What are the largest known structures in the Universe?
12. Where are you in the Universe? If you had to give directions to your location in the Universe, what directions would you give?
13. Approximately what fraction of the age of the Solar System is your life span? Of the age of the Universe?
14. Why should you study astronomy? Do you anticipate needing to know astronomy 5 or 10 years from now? If so, where?
15. How does astronomy help answer the question, "What are we?"
16. **How do we know?** How does the scientific method give scientists a way to know about nature?

ACTIVE INQUIRY QUESTIONS

1. You and three of your friends have won an all-expenses-paid, one-time-only, round-trip first-class vacation to anywhere in the Universe, so long as the choice of destination is unanimous. Where do you want to go, and how do you convince your friends to agree?
2. Think back to the last time you got a new phone and had to figure out how it worked. In what way did you employ the scientific method, maybe without even realizing that you did?

PROBLEMS

(Give your answers in scientific notation when appropriate.)

1. The equatorial diameter of Earth is 7928 miles. If a mile equals 1.609 km, what is Earth's diameter in kilometers? In centimeters?
2. The equatorial diameter of the Moon is 3476 kilometers. If a kilometer equals 0.6214 miles, what is the Moon's diameter in miles?
3. One astronomical unit (AU) is about 1.5×10^8 km. Explain why this is the same as 150×10^6 km.
4. A typical galaxy is shown on the first page of the Universe Bowl on the inside cover of the printed book. Express the number of stars in this typical galaxy in scientific notation.
5. The time of the Cambrian explosion is listed on the second page of the Universe Bowl on the inside cover of the printed book. Express that time in scientific notation.
6. Venus orbits 0.72 AU from the Sun. What is that distance in kilometers? (*Hint*: See Problem 3.)
7. Light from the Sun takes 8 minutes to reach Earth. How long does it take to reach Mars, 1.5 AU from the Sun?
8. The Sun is almost 400 times farther from Earth than is the Moon. How long does light from the Moon take to reach Earth?
9. If the speed of light is 3.0×10^5 km/s, how many kilometers are in a light-year? How many meters? (*Hint*: First look up or calculate how many seconds are in a year.)
10. Light from the star Betelgeuse takes 640 years to reach Earth. How far away is Betelgeuse in units of light-years? Name any historical event that was occurring on Earth at about the time the light left Betelgeuse. Is the distance to Betelgeuse unusual compared with other stars?
11. How long does it take light to cross the diameter of the Milky Way Galaxy?
12. The nearest galaxy to our home galaxy is about 2.5 million light-years away. How many meters is that?
13. How many galaxies like our own would it take if they were placed edge-to-edge to reach the nearest galaxy? (*Hint*: See Problems 11 and 12.)

SENSE OF PROPORTION

1. Arrange the following in order of increasing size: Earth; the Sun; Earth's orbit around the Sun; the Moon; you; the Milky Way Galaxy; the Moon's orbit around Earth
2. Arrange the following in order of increasing distance from you: the Sun; Neptune; the Moon; the center of Earth
3. The Sun is roughly 100 times the diameter of Earth. Which of the following distances is roughly 100 times the diameter of the Sun? radius of the Milky Way Galaxy; radius of the Moon; radius of Earth's orbit around the Sun; radius of the Moon's orbit around Earth
4. The Sun is roughly 100 times the diameter of Earth. If you built a model solar system with a golf ball (diameter 1.68 inches) representing Earth, how big would the model Sun be? Can you think of a common object about that size?

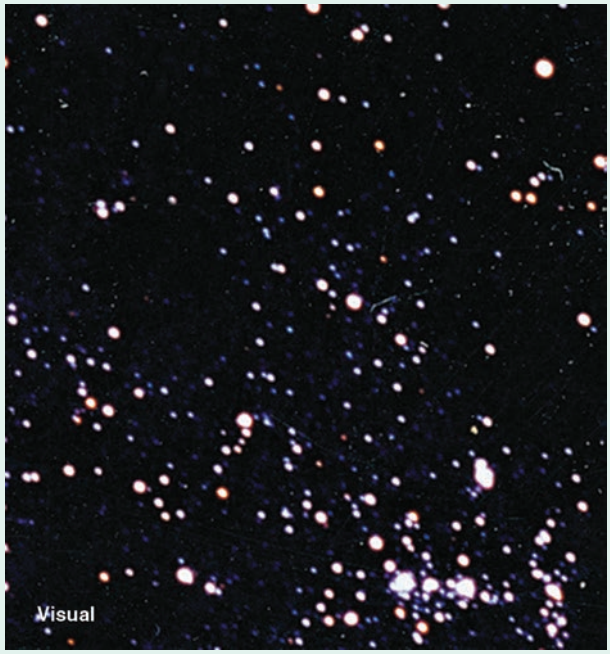
LEARNING TO LOOK

1. Look at the center of Figure 1-4. Approximately what time of day is it at that location? Sunrise? Sunset? Noontime? Midnight? How do you know?
2. Look at Figure 1-6. How can you tell that Mercury does not follow a circular orbit?
3. Look at Figure 1-9. How many stars are within 5 ly of the Sun? Would that number be about the same or much different if Earth orbited another star in this figure?
4. Look at Figure 1-9. Would you say that the distribution of stars is uniform in this field of view, 17 ly across? Compare with Figure 1-10, 1700 ly across. Now look at Figure 1-12; would you say that the distribution of galaxies is uniform in this field of view, 17 million ly across? Compare with Figure 1-13, 1.7 billion ly across.
5. Of the objects listed here, which would be contained inside the object shown in the accompanying photograph? Which would contain the object in the photo?

   star
   planet
   galaxy cluster
   supercluster
   filament
   spiral arm

NOAO/AURA/NSF/Bill Schoening

6. In the photograph shown here, which stars are brightest, and which are faintest? How can you tell? Why can't you tell which stars in this photograph are biggest or which have planets?



NSF/AURA/NOAO

A User's Guide to the Sky

2

Peter Burnett/E+/Getty Images

▲ The sky belongs to everyone equally, and you can enjoy its splendors as if you owned them.

The previous chapter took you on a cosmic zoom through space and time. That quick preview prepared you for the journey to come. In this chapter you can begin your exploration by viewing the sky from Earth; as you do, consider five important questions:

- How are stars and constellations named?
- How are the brightnesses of stars measured and compared?
- How does the sky appear to change and move in daily and annual cycles?
- What causes seasons?
- How do astronomical cycles affect Earth's climate?

As you read about the sky and its motions, notice that the words often seem to imply that Earth is stationary at the center of the Universe. Remind yourself that Earth is really a planet spinning on an axis and moving in an orbit. The next chapter will introduce you to other impressive sky phenomena: phases of the Moon and eclipses.

> **The Southern Cross I saw every night abeam. The sun every morning came up astern; every evening it went down ahead. I wished for no other compass to guide me, for these were true.**
>
> CAPTAIN JOSHUA SLOCUM
> *SAILING ALONE AROUND THE WORLD*

THE NIGHT SKY is the rest of the Universe as seen from Earth. When you look up at the stars, you are looking out through a layer of air only about 100 kilometers (60 miles) deep. Beyond that, space is nearly empty, with the planets of our Solar System several astronomical units away and the distant stars scattered many light-years apart.

As you read this chapter, you will learn about how Earth's motions affect what you can see from your planet, a moving platform:

- Because Earth rotates on its axis once a day, the sky appears to turn around you in a daily cycle. Not only does the Sun rise in the eastern part of the sky and set in the western part, but so do the stars and other celestial objects.
- Because Earth revolves around the Sun once a year, different stars are visible in the night sky in an annual cycle.
- The sequence of seasons you experience is caused by a combination of Earth's yearly motion plus the tilt of Earth's axis relative to its orbit.

2-1 Stars and Constellations

On a dark night far from city lights, you can see a few thousand stars. Long ago, humans tried to make sense of what they saw by naming stars and groups of stars. Some of those ancient names are still in use today.

2-1a Constellations

All around the world, native cultures celebrated heroes, gods, and mythical beasts by giving their names to groups of stars—**constellations** (Figure 2-1). The constellations named within Western culture originated in the civilizations of Assyria, Babylon, Egypt, and Greece more than 3000 years ago. Some constellations look like their namesakes, but others don't, for reasons lost in the passage of millennia.

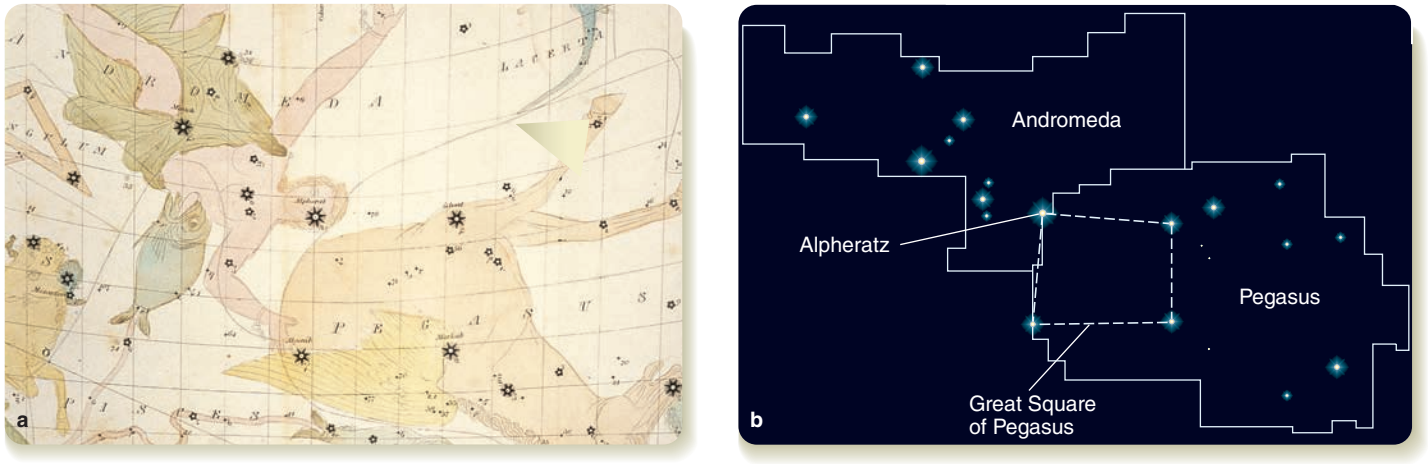
Different cultures grouped stars and named constellations differently. The constellation you call Orion was known in antiquity as Al Jabbār (the Giant) to the Arabs, as the White Tiger to the Chinese, and as Prajapati (a deity in the form of a stag) in India. The Pawnee Indians saw the constellation Scorpius as two groupings: The long tail of the scorpion was the Snake, and the bright stars at the tip of the scorpion's tail were the Two Swimming Ducks. In Hawai'i, the scorpion's tail was Maui's Fishhook that pulled the islands up from the bottom of the ocean.

On the other hand, many cultures, including the ancient Greeks, northern Asians, and Native Americans, all associated the stars in and around the Big Dipper with the figure of a bear. The concept of the celestial bear may have crossed the land bridge into North America with the first Americans more than 12,000 years ago. Hence, the names of some of the groups of stars you see in the sky may be among the oldest surviving traces of human culture.

Originally, constellations were simply loosely defined grouping of bright stars. Many of the fainter stars were not included in any constellation, and stars in the southern sky, not visible to early astronomers observing from northern latitudes, were not included on their star maps. Constellation boundaries, when they were defined at all, were only

◀ **Figure 2-1** The constellations are an ancient heritage handed down for thousands of years as celebrations of mythical heroes and monsters. Here, Sagittarius and Scorpio appear above the southern horizon.

From Duncan Bradford, Wonders of the Heavens(Boston: John B. Russell, 1837). © 2016 Cengage

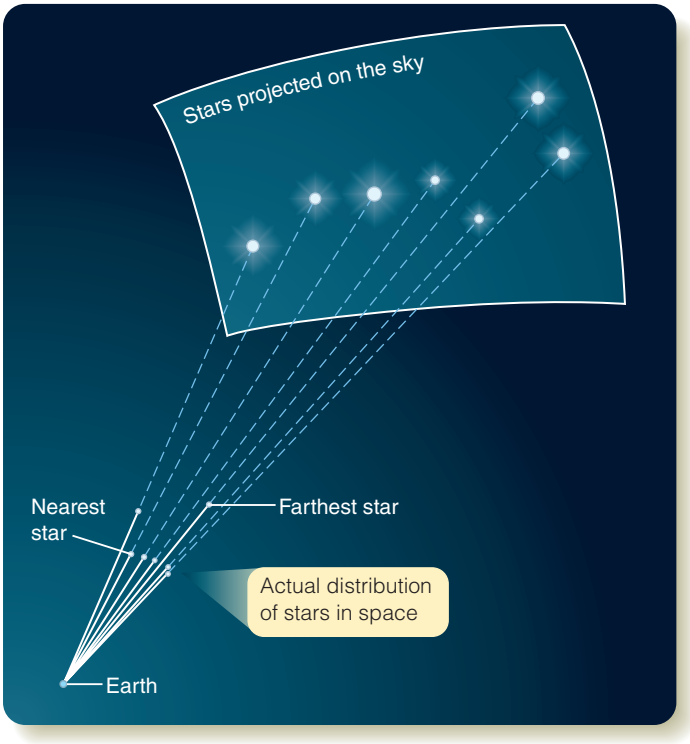


▲ **Figure 2-2** (a) In antiquity, constellation boundaries were poorly defined, as shown on this map by the curving dotted lines separating Pegasus from Andromeda. (b) Modern constellation boundaries were precisely defined by international agreement in 1928.

approximate (**Figure 2-2a**), so a star like Alpheratz could be thought of as both part of Pegasus and part of Andromeda. To correct these gaps and ambiguities, modern astronomers invented more constellations, and in 1928 the **International Astronomical Union (IAU)** established 88 official constellations with carefully defined boundaries (**Figure 2-2b**) that together include every part of the sky. (The IAU is the same organization that redefined Pluto as a dwarf planet in 2006, as mentioned in Chapter 1.) Consequently, the modern scientific use of constellations is to designate precisely defined sections of the sky, such that any star within the area belongs to just that one constellation. Now, Alpheratz is only in Andromeda.

In addition to the 88 official constellations, the sky contains a number of less formally defined groupings called **asterisms**. The Big Dipper, for example, is a well-known asterism that is part of the constellation Ursa Major (the Great Bear). Another asterism is the Great Square of Pegasus (Figure 2-2b) that includes three stars from Pegasus plus Alpheratz from Andromeda. You can introduce yourself to the brighter constellations and asterisms using the star charts in Appendix B.

Although constellations and asterisms are groups of stars that appear close together in the sky, it is important to remember that most are made up of stars that are not physically associated with one another. Some stars may be many times farther away than others and moving through space in different directions. The only thing they have in common is that they happen to lie in approximately the same direction as seen from Earth (**Figure 2-3**).



▲ **Figure 2-3** When you see the Big Dipper in the sky, you are actually looking through a group of stars scattered through space at different distances from Earth. You view them as if they were projected on a screen, and they form the shape of the Dipper.

2-1b Star Names

In addition to naming groups of stars, early astronomers gave individual names to the brightest stars. Modern astronomers still use many of those ancient names. Although the constellation names came from Greek translated into Latin—the languages of science until the 19th century—most individual star names come from Arabic and have been altered through the passing centuries. The name of Betelgeuse, the bright orange star in Orion, for example, comes from the Arabic *yad al-jawza*, meaning "Hand of Jawza [Gemini and Orion]." Names such as Sirius (Scorcher) and Aldebaran (the Follower [of the Pleiades]) are intriguing additions to the mythology of the sky.

Naming individual stars is not very helpful because you can see thousands of them. How many names could you possibly remember? Also, a simple name gives you little or no information about the star itself. A more useful way to identify stars was invented by German astronomer Johann Bayer in a catalog published in 1603. Bayer assigned Greek letters to the bright stars in each constellation in an order determined by a combination of brightness and position. Astronomers still use Bayer's designations for bright stars. If a star's label is early in the Greek alphabet—for instance Alpha or Beta—you can be sure that star is among the brightest stars in that constellation. Often the name of the Greek letter is spelled out, as in "Alpha," but sometimes the actual Greek letter is used, especially in charts. (You can find the Greek alphabet in Appendix A.) The Bayer designations for the stars in Orion are shown in **Figure 2-4**.

To identify a star by its Greek-letter designation, you would give the Greek letter followed by the genitive (possessive) form of the constellation name; for example, the brightest star in the constellation Canis Major is Alpha Canis Majoris, which can also be written as α Canis Majoris. This name identifies the star and the constellation and gives a clue to the star's relative brightness. Compare this with the ancient "personal" name for this star, Sirius, which tells you nothing about its location or brightness.

2-1c Famous Stars

It is fun to know the names of the brighter stars, but they are more than points of light in the sky. Each is a glowing sphere of gas resembling the Sun, with its own unique characteristics. **Figure 2-5** identifies eight bright stars that can be considered as Famous Stars. As you read this book, you will encounter each of them several more times, discover their colorful personalities, and enjoy finding them in the evening sky. You will learn, for example, that Betelgeuse is not just an orange point of light but is an aging, cool star more than 500 times larger than the Sun. As you explore further in later chapters, you may want to nominate more Famous Stars to the list.

You can use the star charts at the end of this book to help you locate these Famous Stars. You can see Polaris throughout the year from the Northern Hemisphere, but Sirius, Betelgeuse, Rigel, and Aldebaran are only in the winter sky. Spica is a summer star, and Vega is visible on summer and fall evenings. Alpha Centauri, only 4.4 ly away, is the bright star that is nearest to us, but you have to travel to the latitude of southern Florida to glimpse it above the southern horizon.

2-1d Star Brightness

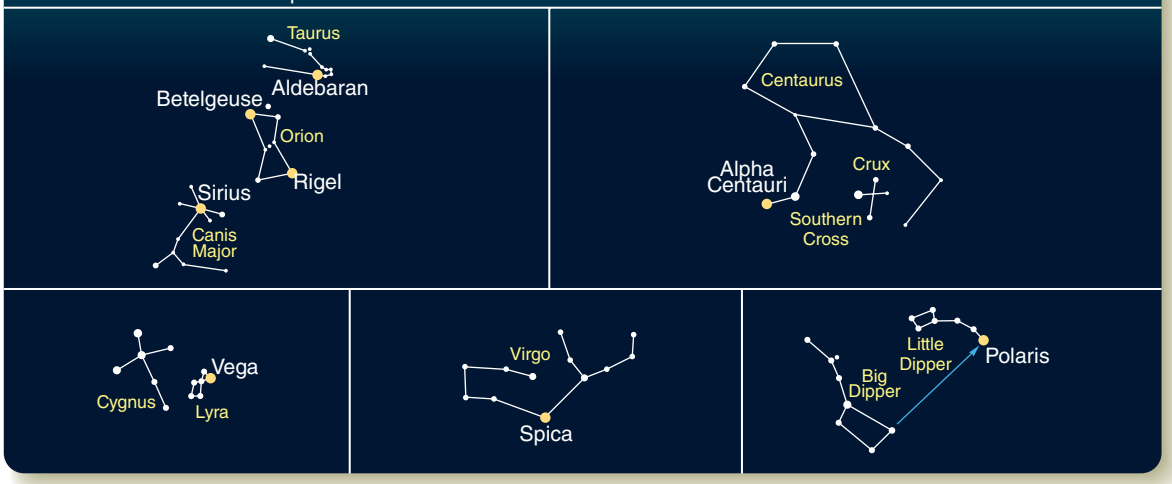
Astronomers describe the brightness of stars using the **magnitude scale**, a system that first appeared in the writings of the astronomer Claudius Ptolemaeus (pronounced TAHL-eh-MAY-us) about the year 140. The magnitude system probably originated even earlier; many historians attribute its invention to the Greek astronomer Hipparchus (about 190–120 BCE) who compiled the first known star catalog. Almost 300 years later, Ptolemy used the magnitude scale in his catalog, which was substantially based on Hipparchus's previous work, and successive generations of astronomers have continued to use their system.

Those early astronomers divided the stars into six classes. The brightest stars were called first-magnitude stars and the

William Hartmann

◀ **Figure 2-4** Greek letter designations for the stars of Orion: α (Alpha) and β (Beta) are the two brightest stars in the constellation, although Alpha is variable and usually fainter than Beta. Fainter stars do not have Greek letters or names, but if they are located inside the constellation boundaries, they are part of the constellation. The brighter stars in a constellation often also have individual names derived from Arabic. (The spikes on the star images in the photograph were produced by an optical effect in the telescope.)

Star Common Name	Bayer Designation	Special Feature	Visible in the Northern Hemisphere
Betelgeuse	Alpha Orionis	Bright red star in Orion	Winter
Aldebaran	Alpha Tauri	Red eye of Taurus the Bull	Winter
Sirius	Alpha Canis Majoris	Brightest star in the sky	Winter
Rigel	Beta Orionis	Bright blue star in Orion	Winter
Rigil Kentaurus	Alpha Centauri	Nearest bright star to the Sun	Spring, far south
Vega	Alpha Lyrae	Brightest star in summer sky	Summer
Spica	Alpha Virginis	Bright southern star	Summer
Polaris	Alpha Ursae Minoris	The North Star	Year round



◀ **Figure 2-5** Famous Stars: Locate these bright stars in the sky and learn about their characteristics. Refer to the star charts in Appendix B.

next brightest set, second-magnitude stars. The scale continued downward to sixth-magnitude stars, the faintest visible to the human eye. Thus, the larger the magnitude number, the fainter the star. This might make sense if you think of the brightest stars as first-class stars and the faintest visible stars as sixth-class stars.

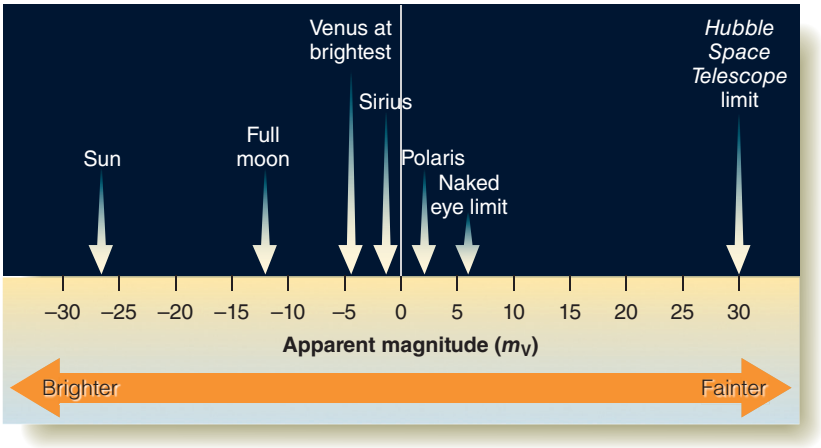
Ancient astronomers could only estimate magnitudes by eye, but modern astronomers can use scientific instruments to measure the brightness of stars to high precision, so they have carefully redefined the magnitude scale. For example, instead of saying that the star known by the charming name Chort (Theta Leonis) is third magnitude, they can say its magnitude is 3.34. In the redefined scale, some stars are actually brighter than magnitude 1.0. For example, Famous Star Vega (also known as Alpha Lyrae) is so bright that its magnitude, 0.03, is almost zero. A few are so bright that the modern magnitude scale must extend into negative numbers (**Figure 2-6**). On this scale, Famous Star Sirius—the brightest star in the sky—has a magnitude of –1.46. Modern astronomers have had to extend the faint end of the magnitude scale as well. The faintest stars you can see with your unaided eyes are about sixth magnitude, but if you use a telescope, you can detect stars much fainter than that. Magnitude numbers larger than 6 are needed to describe such faint stars.

These numbers are known as **apparent visual magnitudes** (m_V) because they describe how the stars look to human eyes observing from Earth. Although some stars emit relatively large amounts of infrared or ultraviolet light, human eyes can't see those types of radiation, and they are not included in the apparent visual magnitude. The subscript V stands for *visual* and reminds you that only visible light is included. Also, apparent visual magnitude does not take into account the distance to the stars. In other words, a star's apparent visual magnitude tells you only how bright the star looks as seen from Earth, not about its actual light output.

2-1e Magnitude and Flux

Your interpretation of brightness is quite subjective, depending on both the physiology of human eyes and the psychology of perception. As a careful investigator, you should refer to **flux**, which is a measure of the light energy from a star that hits a collecting area of one square meter in one second. Such measurements precisely and objectively define the brightness of starlight.

Astronomers use a simple formula to convert between magnitudes and flux. If two stars have fluxes F_A and F_B, then the ratio of their fluxes is F_A/F_B. To make today's measurements agree with ancient catalogs, astronomers have defined the modern magnitude scale so that two stars differing in brightness by 5 magnitudes have a flux ratio of exactly 100. Therefore, two stars that differ by 1 magnitude must have a flux ratio that equals the fifth root of



◀ **Figure 2-6** The scale of apparent visual magnitudes extends into negative numbers to represent the brightest objects and into positive numbers larger than six to represent objects fainter than the unaided human eye can see.

100, symbolized by $\sqrt[5]{100}$ or $100^{0.2}$, which is about 2.51; that is, the light from one star must be approximately 2.51 times brighter (has 2.51 times more flux arriving at Earth) than the other.

You can practice using this definition for other pairs of stars. For example, if two stars differ in brightness by 3 magnitudes they will have a flux ratio of approximately 2.51 × 2.51 × 2.51, which is 2.51^3 or about 15.8. **Table 2-1** shows the flux ratios corresponding to various magnitude differences. For example, suppose one star is third magnitude and another star is ninth magnitude. What is their flux ratio? In this case, the magnitude difference is six, and Table 2-1 shows the equivalent flux ratio is about 251. Therefore, light from one star is about 251 times brighter than light from the other star.

A table is convenient, but for more precision you can use the relationship expressed as a simple formula. The flux ratio F_A/F_B is equal to 2.51 raised to the power of the magnitude difference $m_B - m_A$:

KEY EQUATION 2-1a Magnitude and Flux

$$\frac{F_A}{F_B} = (2.51)^{(m_B - m_A)}$$

Example 1: If the magnitude difference between two stars is 6.32, what is the flux ratio? ***Solution:*** The flux ratio must be $2.51^{6.32}$. A calculator tells you the answer: 336. Star A is about 336 times brighter than Star B, meaning that the flux received on Earth from Star A is 336 times greater than that from Star B.

On the other hand, if you know the flux ratio of two stars and want to find their magnitude difference, it is convenient to rearrange the preceding formula and write it as:

KEY EQUATION 2-1b Magnitude and Flux

$$m_B - m_A = 2.5 \log\left(\frac{F_A}{F_B}\right)$$

The expression *log* means logarithm to the base 10. ***Example 2:*** The light from Sirius has 24.2 times as much flux (is 24.2 times brighter) than light from Polaris. What is their magnitude difference? ***Solution:*** The magnitude difference is therefore 2.5 log (24.2). Your pocket calculator tells you the logarithm of 24.2 is 1.384, so the magnitude difference is 2.5 × 1.384 which equals 3.46 magnitudes. Thus, Sirius is 3.46 magnitudes brighter than Polaris.

The modern magnitude system, although seemingly complicated, has some advantages. It compresses a tremendous range of brightness into a small range of magnitudes, as you can see in Table 2-1. More important, it allows modern astronomers to measure and report the brightness of stars to high precision while

Table 2-1 Magnitude Differences and Flux Ratios

Magnitude Difference	Corresponding Flux Ratio
0.00	1.00
1.00	2.51
2.00	6.31
3.00	15.8
4.00	39.8
5.00	100
6.00	251
7.00	631
8.00	1580
9.00	3980
10.0	10,000
:	:
:	:
15.0	1,000,000
20.0	100,000,000
25.0	10,000,000,000
:	:
:	:

HOW DO WE KNOW? 2-1

Scientific Models

How can a scientific model be useful if it is not entirely true?

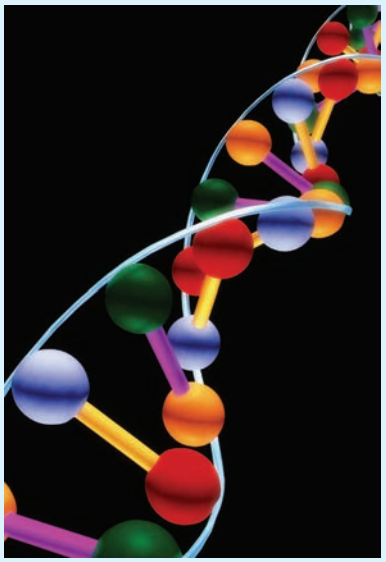
A scientific model is a carefully devised conception of how something works; that is, a framework that helps scientists think about some aspect of nature, just as the celestial sphere helps astronomers think about the motions of the sky.

Chemists, for example, use colored balls to represent atoms and sticks to represent the bonds between them, kind of like Tinkertoys. Using these molecular models, chemists can see the three-dimensional shape of molecules and understand how the atoms interconnect.

You have probably seen elaborate ball-and-stick models of DNA, but does the molecule really look like Tinkertoys? No, but the model is both simple enough and accurate enough to help scientists think productively about the molecule. The molecular model of DNA proposed by James D. Watson and Francis Crick in 1953 led to our modern understanding of the mechanisms of genetics.

A scientific model is not a statement of truth; it does not have to be precisely true to be useful. In an idealized model, some complex aspects of nature can be simplified or omitted. The ball-and-stick model of a molecule doesn't show the relative strength of the chemical bonds, for instance. A model gives scientists a way to think about some aspect of nature but need not be true in every detail.

When you use a scientific model, it is important to remember the limitations of that model. If you begin to think of a model as true, it can be misleading instead of helpful. The celestial sphere, for instance, can help you think about the sky, but you must remember that it is only a model. The Universe is much larger and much more interesting than this early scientific model of the heavens.



Alfred Pasieka/Science Source

Balls represent atoms and rods represent chemical bonds in this model of a DNA molecule.

remaining connected to observations of apparent visual magnitude that go back to the time of Hipparchus.

2-2 The Sky and Celestial Motions

The sky above seems like a great blue dome in the daytime and a sparkling ceiling at night. It was this domed ceiling that the first astronomers had in mind long ago as they tried to understand the Universe.

2-2a The Celestial Sphere

Ancient astronomers believed the sky was a great sphere surrounding Earth with the stars stuck on the inside like thumbtacks. Modern astronomers know that the stars are scattered through space at different distances, but it is still convenient to think of the sky as a great starry sphere enclosing Earth.

The celestial sphere is an example of a **scientific model**, a common feature of scientific thought (**How Do We Know? 2-1**). Notice that a scientific model does not have to be true to be useful. You will encounter many scientific models in the chapters that follow, and you will discover that some of the most useful models are highly simplified descriptions of the true facts.

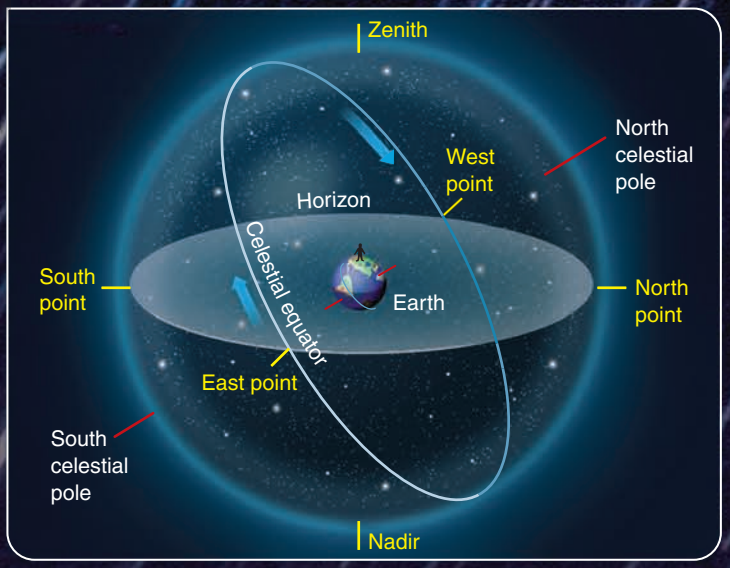
Concept Art: **The Sky Around You** takes you on an illustrated tour of the sky. Throughout this book, these Concept Art pages introduce new concepts and new terms through photos and diagrams, so be sure to examine them carefully. **The Sky Around You** introduces you to three important principles and 16 new terms that will help you understand the sky:

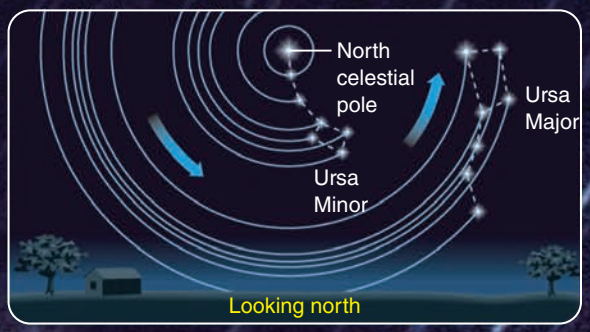
1. The sky appears to rotate westward around Earth each day, but that is a consequence of the eastward rotation of Earth. That rotation produces day and night. Notice how reference points on the *celestial sphere* such as the *zenith, nadir, horizon, celestial equator, north celestial pole,* and *south celestial pole* define the four cardinal directions, *north point, south point, east point,* and *west point.*
2. Astronomers measure *angular distance* across the sky as angles and express them as degrees, *arc minutes*, and *arc seconds.* The same units are used to measure the *angular diameter* of an object.
3. What you can see of the sky depends on where you are on Earth. If you live in Australia, you can see many stars, constellations, and asterisms invisible from North America, but you would never see the Big Dipper.

The Sky Around You

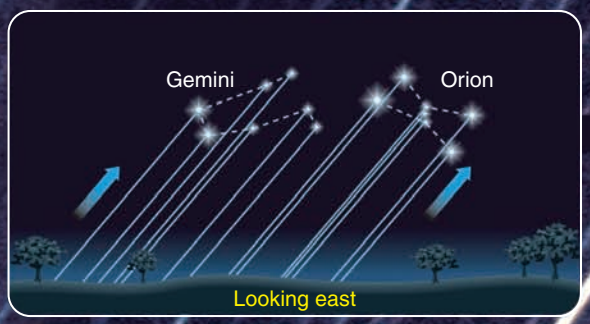
1 The eastward rotation of Earth causes the Sun, Moon, planets, and stars to move westward in the sky as if the **celestial sphere** were rotating westward around Earth. From any location on Earth you see only half of the celestial sphere, the half above the **horizon**. The **zenith** marks the point of the celestial sphere directly above your head, and the **nadir** marks the point of the celestial sphere directly under your feet. The drawing at right shows the view for an observer in North America. An observer in South America would have a completely different horizon, zenith, and nadir.

The apparent pivot points are the **north celestial pole** and the **south celestial pole** located directly above Earth's north and south poles. Halfway between the celestial poles lies the **celestial equator**. Earth's rotation defines the directions you use every day: the **north point** and **south point** are the points on the horizon closest to the celestial poles, and the **east point** and the **west point** lie halfway between the north and south points. The celestial equator always meets the horizon at the east and west points.

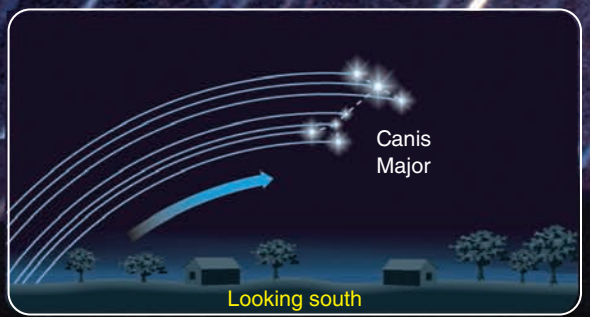




Looking north, you see Famous Star Polaris (the North Star) located near the north celestial pole. As the sky appears to rotate westward, Polaris hardly moves, but other stars circle the celestial pole.



Looking east, you see stars rising on tracks that are small portions of large circles extending around the sky.

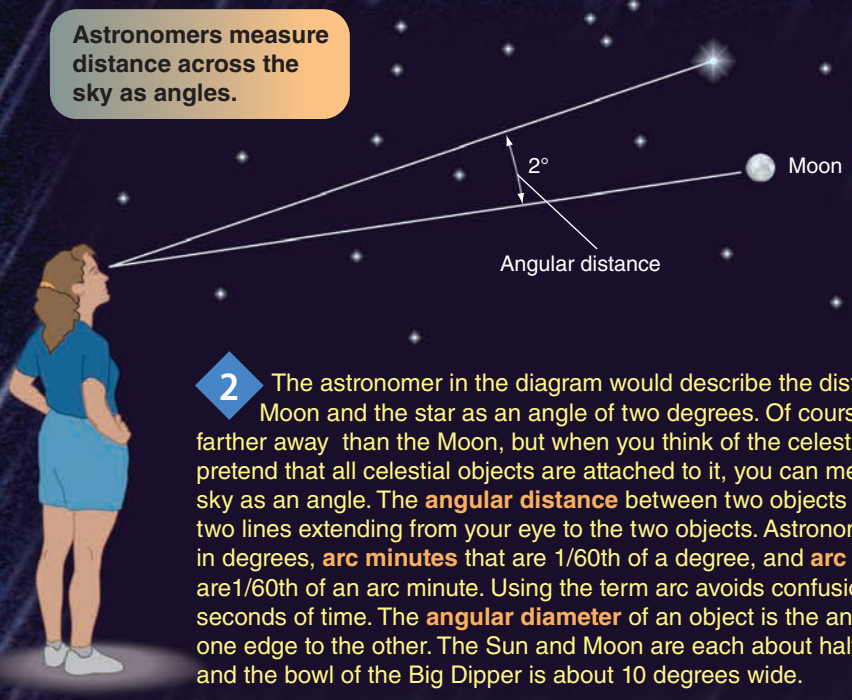


Looking south from a location in North America, you

1a This time exposure of about 30 minutes shows stars as streaks, called star trails, rising behind an observatory dome lit from below by red night lights. The camera was facing northeast to take this photo. The motion you see in the sky depends on which direction you look, as shown in the three inset diagrams.

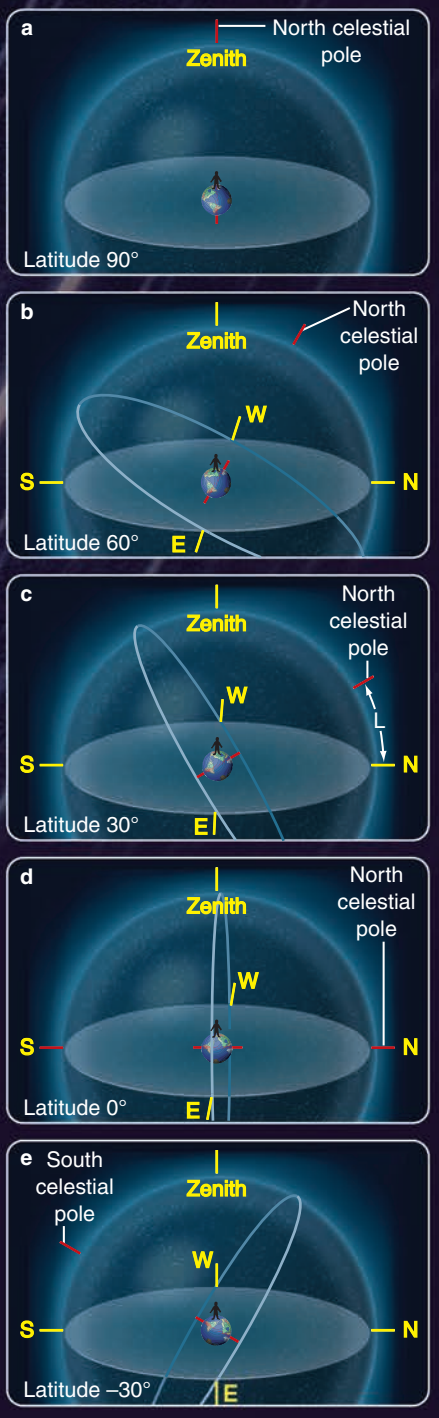
NSF/AURA/NOAO; © 2016 Cengage

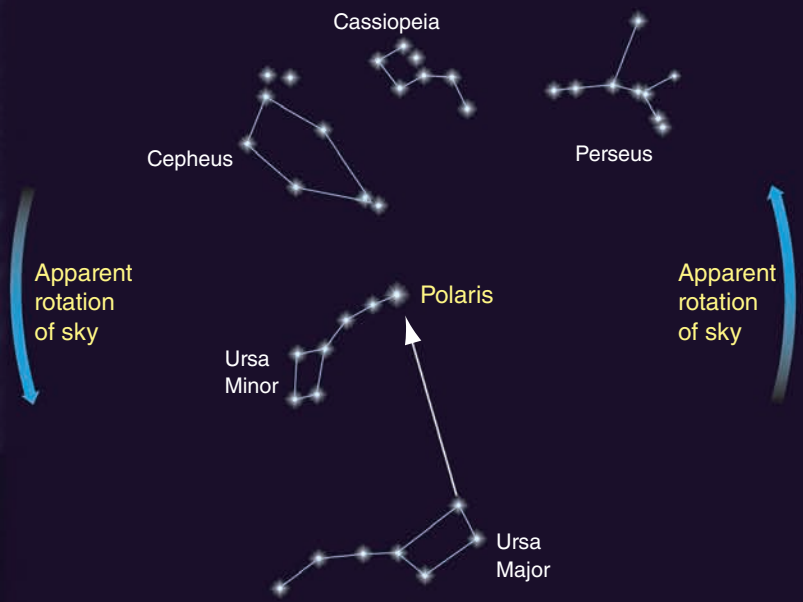
Astronomers measure distance across the sky as angles.



2 The astronomer in the diagram would describe the distance between the Moon and the star as an angle of two degrees. Of course, the star is much farther away than the Moon, but when you think of the celestial sphere, and pretend that all celestial objects are attached to it, you can measure distance on the sky as an angle. The **angular distance** between two objects is the angle between two lines extending from your eye to the two objects. Astronomers measure angles in degrees, **arc minutes** that are 1/60th of a degree, and **arc seconds** that are1/60th of an arc minute. Using the term arc avoids confusion with minutes and seconds of time. The **angular diameter** of an object is the angular distance from one edge to the other. The Sun and Moon are each about half a degree in diameter, and the bowl of the Big Dipper is about 10 degrees wide.

3 What you see in the sky depends on your latitude, as shown at right. Imagine that you begin a journey at Earth's North Pole with the north celestial pole directly overhead (a). As you walk southward, the celestial pole moves toward the horizon, and you can see further into the southern sky (b). The angular distance (L) from the horizon to the north celestial pole shown in part c always equals your latitude—an important basis for celestial navigation. As you cross Earth's equator (d), the celestial equator would pass through your zenith, and the north celestial pole would sink below your northern horizon (e).





3a **Circumpolar constellations** are those that never rise or set because they are close to the celestial poles. From mid-northern latitudes, as shown at left, you see a number of familiar constellations circling Polaris and never dipping below the horizon. As Earth turns and the sky appears to rotate, the pointer stars at the front of the Big Dipper always point approximately toward Polaris. Circumpolar constellations near the south celestial pole never rise as seen from mid-northern latitudes. From a high northern latitude location such as Norway (b), you would have more circumpolar constellations, and from Quito, Ecuador, located on Earth's equator (d), you would have no circumpolar constellations at all.

AURA/NOAO/NSF; © Cengage

How many *circumpolar constellations* you see depends on where you are. Remember Famous Star Alpha Centauri? It is in the southern sky and is not visible from most of the United States, but you can see it easily from Australia.

Pay special attention to these new terms. You need to know these terms to describe the sky and celestial motions, but don't fall into the trap of just memorizing new terms. The goal of science is to understand nature, not to memorize definitions. Study the diagrams and see how the geometry of the celestial sphere and its apparent motions explain the changing appearance of the sky above you.

This is a good time to consider a couple of **Common Misconceptions**. Many people, without thinking about it much, assume that the stars are not in the sky during the daytime. The stars are actually there day and night; they are just invisible during the day because the sky is lit by the Sun. Also, many people insist that Famous Star Polaris is the brightest star in the sky. It is actually the 50th visually brightest star. Polaris is important because of its position, not because of its brightness.

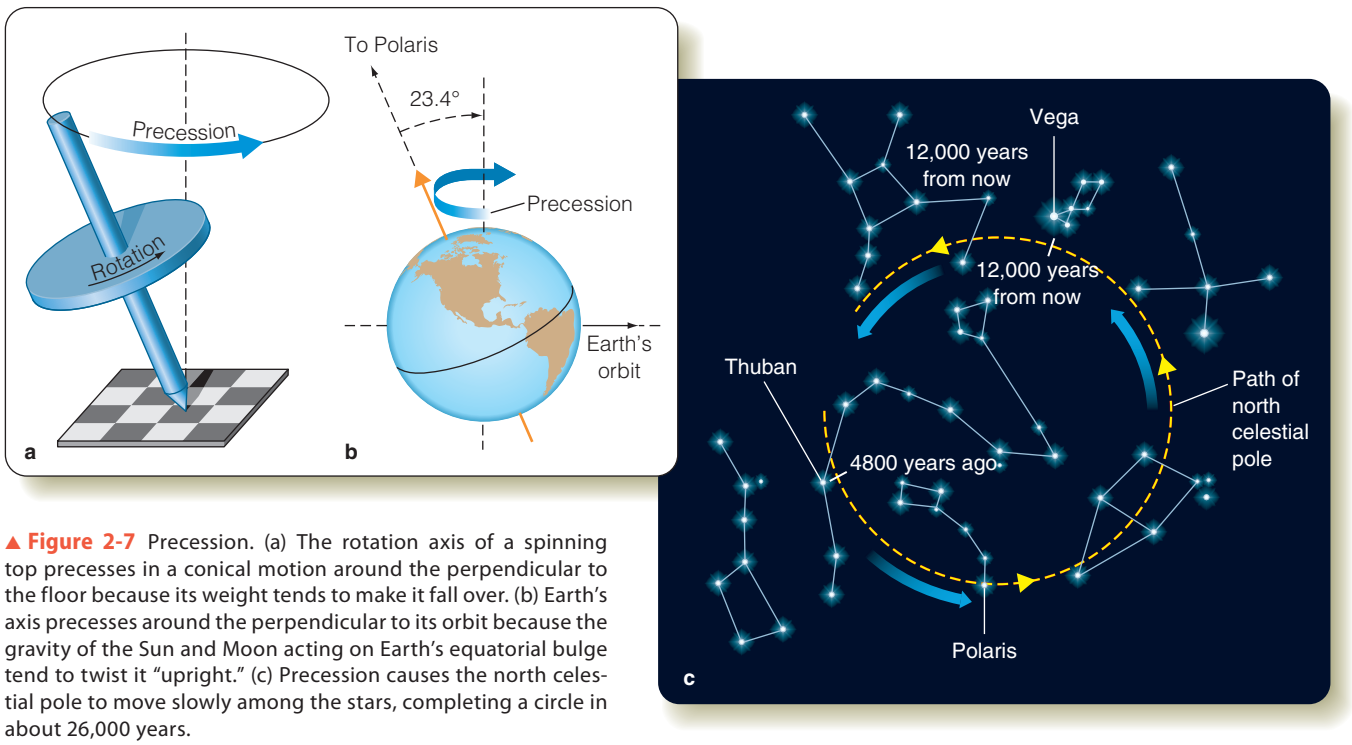
2-2b Precession

In addition to causing the obvious daily motion of the sky, Earth's rotation is connected with a very slow celestial motion that can be detected only over centuries. More than 2000 years ago, Hipparchus compared positions of some stars with their positions recorded nearly two centuries earlier and realized that the celestial poles and equator were slowly moving across the sky. Later astronomers understood that this motion is caused by a toplike motion of Earth known as **precession**.

If you have ever played with a gyroscope or top, you have seen how the spinning mass resists any sudden change in the direction of its axis of rotation. The more massive the top and the more rapidly it spins, the more it resists your efforts to twist it out of position. You may recall that even the most rapidly spinning top slowly swings its axis around in a circle. The weight of the top tends to make it tip over, and this combines with its rapid rotation to make its axis sweep out the shape of a cone. That motion is precession (**Figure 2-7a**). In later chapters, you will learn that many celestial bodies precess.

Earth spins like a giant top, but it does not spin upright in its orbit; its axis is tipped 23.4 degrees from vertical. Earth's large mass and rapid rotation keep its axis of rotation pointed toward a spot near the star Polaris, and the axis would remain pointed constantly in that direction except for the effect of precession.

Earth has a slight bulge around its middle because of its rotation. The gravity of the Sun and Moon pull on the bulge, tending to twist Earth's axis "upright" relative to its orbit. If Earth were a perfect sphere, it would not be subjected to this twisting force. Notice that the analogy to a spinning top is not perfect; gravity tends to make a top fall over, but it tends to twist Earth upright. In both cases, the twisting of the axis of rotation combined with the rotation of the object causes precession. The precession of Earth's axis takes about 26,000 years for one cycle (**Figure 2-7b**).



▲ **Figure 2-7** Precession. (a) The rotation axis of a spinning top precesses in a conical motion around the perpendicular to the floor because its weight tends to make it fall over. (b) Earth's axis precesses around the perpendicular to its orbit because the gravity of the Sun and Moon acting on Earth's equatorial bulge tend to twist it "upright." (c) Precession causes the north celestial pole to move slowly among the stars, completing a circle in about 26,000 years.

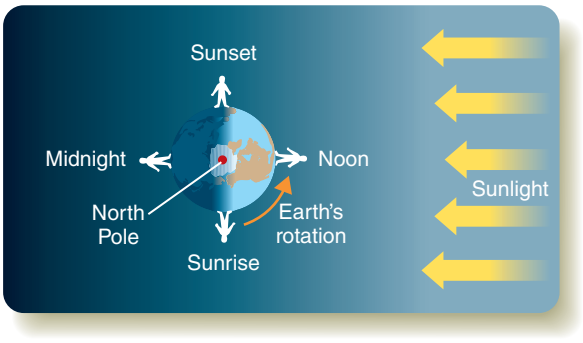
Because the locations of the celestial poles and equator are defined by Earth's rotation axis, precession slowly moves these reference marks. You would notice no change at all from night to night or even year to year, but precise measurements can reveal the slow precession of the celestial poles and the resulting change in orientation of the celestial equator.

Over centuries, precession has significant effects. Egyptian records show that 4800 years ago, the north celestial pole was near the star Thuban (Alpha Draconis). The pole is now moving closer to Polaris and will be closest to it in about the year 2100. In another 12,000 years, the pole will have moved to within 5 degrees of Vega (Alpha Lyrae). Next time you glance at Famous Star Vega, remind yourself that it will someday be an impressive north star. **Figure 2-7c** shows the path through the constellations followed by the north celestial pole during the 26,000-year precession cycle.

2-3 Sun and Planets

Earth's rotation on its axis causes the cycle of day and night, but its motion around the Sun in its orbit defines the year. Notice an important distinction. **Rotation** is the turning of a body on its axis, whereas **revolution** means the motion of a body around a point outside the body. Consequently, astronomers are careful to say Earth rotates once a day on its axis and revolves once a year around the Sun. (This may be difficult to keep straight because it is the opposite of the common English use of those words in relation to automobiles: People normally say that the tires revolve as a car moves, and every once in a while the tires have to be rotated. It would be astronomically correct to say the tires rotate as you drive, and you revolve your tires when they need it, but nobody except an astronomer would know what you meant.)

Because day and night are caused by the rotation of Earth, the time of day depends on your location on Earth. You can notice this if you watch live international news. It may be lunchtime where you are, but for a newscaster in the Middle East, it can already be dark. In **Figure 2-8**, you can see that four people in different places on Earth have different times of day.



▲ **Figure 2-8** This view of Earth as if looking down from above the North Pole shows how the time of day or night depends on your location.

2-3a Annual Motion of the Sun

The sky is filled with stars even in the daytime, but the glare of sunlight fills Earth's atmosphere with scattered light, and you can see only the brilliant Sun. If the Sun were fainter, you would be able to see it rise in the morning with certain stars in its background. During the day you would see the Sun and the stars apparently moving westward, and the Sun would eventually set in front of the same stars it rose with in the morning. If you watched carefully as the day passed, you would notice that the Sun was creeping slowly eastward against the background of stars. It would move a distance roughly equal to its own diameter between sunrise and sunset. This motion is caused by the motion of Earth in its orbit around the Sun.

For example, from mid-December to mid-January, you would see the Sun in front of the constellation Sagittarius (**Figure 2-9**). As Earth moves along its orbit, the Sun appears to move eastward among the stars. By late February, you would see it in front of Aquarius.

Although people say the Sun is "in Sagittarius" or "in Aquarius," it isn't really correct to say the Sun is "in" any constellation. The Sun is only 1 AU away, and the stars visible in the sky are hundreds of thousands or millions of times more distant. Nevertheless, in March of each year, the Sun passes in front of the stars that make up Aquarius, and people conventionally use the expression, "The Sun is in Aquarius."

The apparent path of the Sun against the background of stars is called the **ecliptic**. If the sky were a great screen, the ecliptic would be the shadow cast by Earth's orbit. That is why the ecliptic is often called the *projection* of Earth's orbit on the sky.

Earth circles the Sun relative to the background stars in 365.26 days, and consequently the Sun appears to circle the sky, returning to the same position relative to the background stars, in the same period. That means the Sun, moving 360 degrees around the ecliptic in 365.26 days, travels approximately 1 degree eastward in 24 hours, about twice its angular diameter. You don't notice this apparent motion of the Sun because you can't see the stars in the daytime, but it does have an important consequence that you do notice—the seasons.

2-3b Seasons

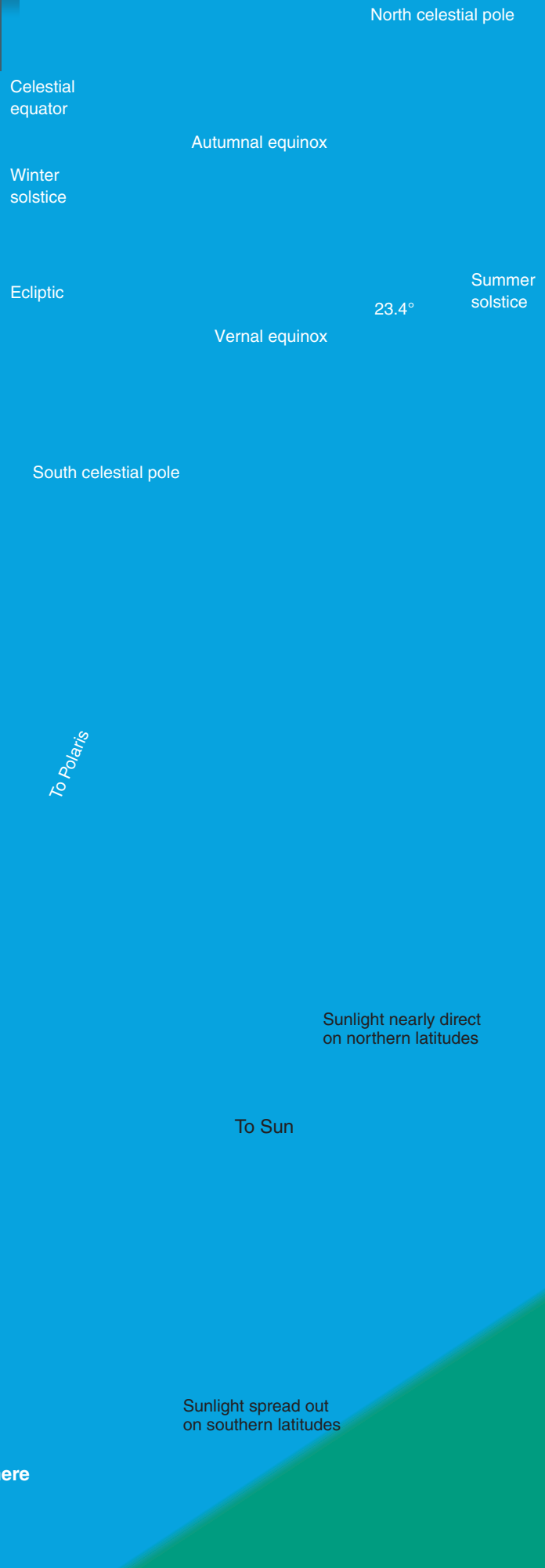
Earth would not experience obvious seasons if it rotated upright in its orbit, but because its axis of rotation is tipped 23.4 degrees from the perpendicular to its orbit, it has seasons. Study **The Cycle of the Seasons** and notice two important principles plus six new terms:

1. Because Earth's axis of rotation is inclined 23.4 degrees, the Sun moves into the northern sky in the spring and into the southern sky in the fall. This is what causes the cycle of the seasons. Notice how the *vernal equinox,* the *summer solstice,* the *autumnal equinox,* and the *winter solstice* mark the beginnings of the seasons. Further, notice the very minor effects of Earth's slightly elliptical orbit as it travels from *perihelion* to *aphelion.*

The Cycle of the Seasons

You can use the celestial sphere to help you think about the seasons. The celestial equator is the projection of Earth's equator on the sky, and the ecliptic is the projection of Earth's orbit on the sky. Because Earth is tipped in its orbit, the ecliptic and equator are inclined to each other by 23.4 degrees, as shown at right. As the Sun moves eastward around the sky, it spends half the year in the southern half of the sky and half the year in the northern half. That causes the seasons.

The Sun crosses the celestial equator going northward at the point called the **vernal equinox**. The Sun is at its farthest north at the point called the **summer solstice**. It crosses the celestial equator going southward at the **autumnal equinox** and reaches its most southern point at the **winter solstice**.

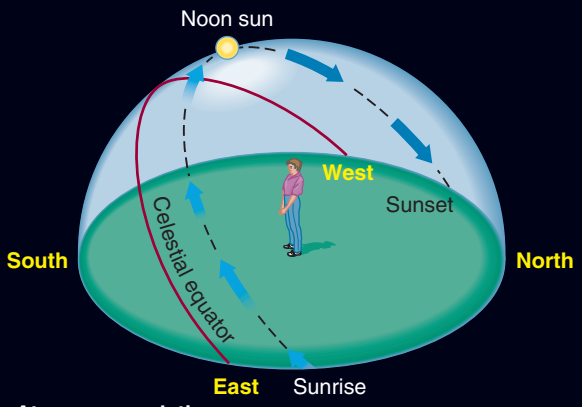


Earth at Northern Hemisphere summer solstice

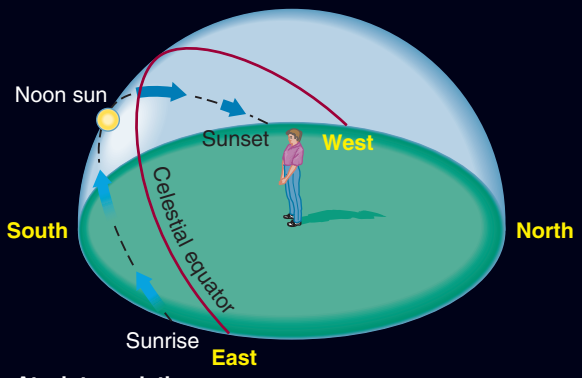
1c Light striking the ground at a steep angle spreads out less than light striking the ground at a shallow angle. Light from the summer solstice Sun strikes northern latitudes from nearly overhead and is concentrated.

Light from the winter solstice sun strikes northern latitudes at a much steeper angle and spreads out. The same amount of energy is spread over a larger area, so the ground receives less energy from the winter sun.

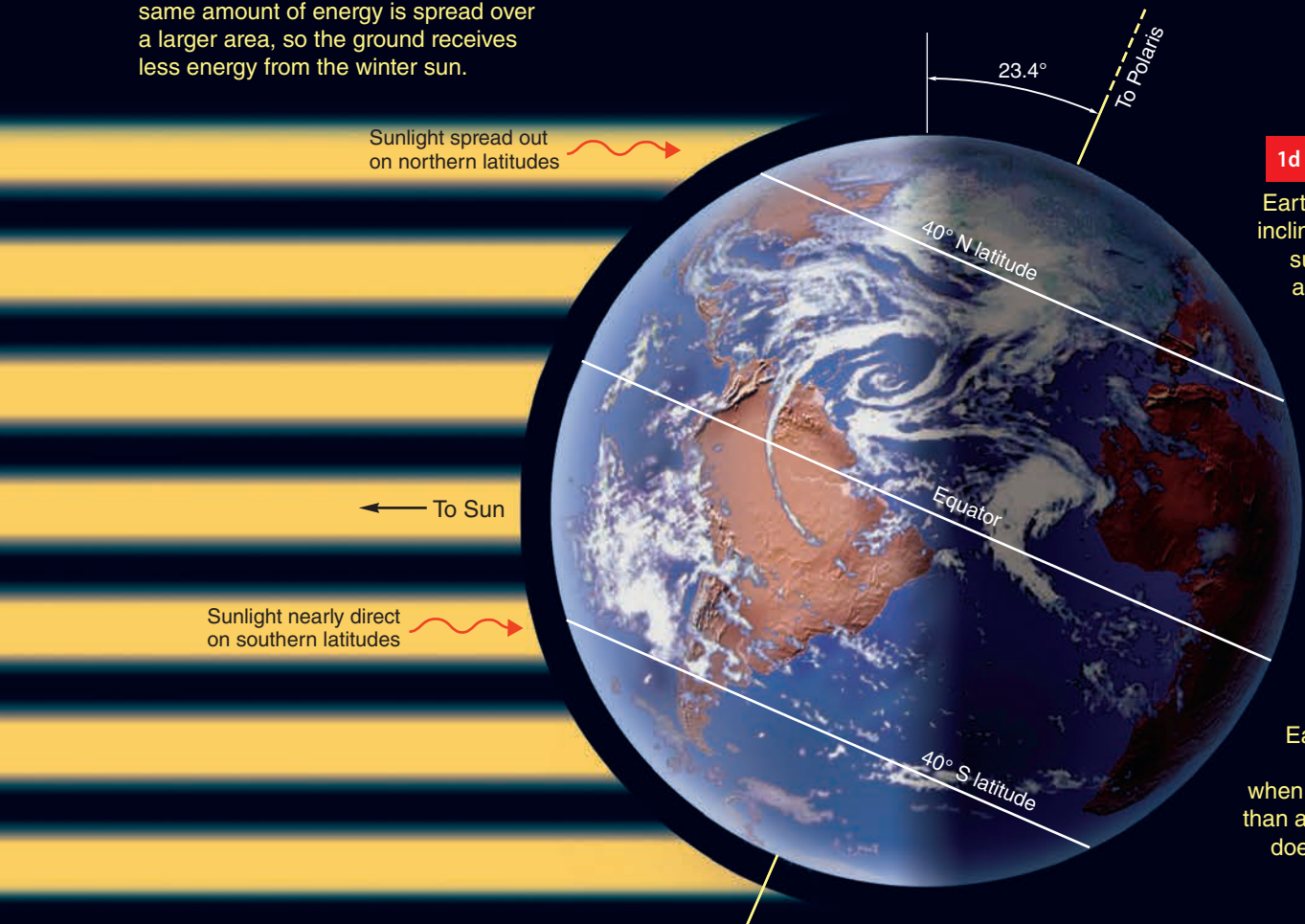
2 The two causes of the seasons in the Northern Hemisphere are shown at right. First, the noon summer Sun is higher in the sky and the winter Sun is lower, as shown by the longer winter shadows. Thus, winter sunlight is more spread out. Second, the summer Sun rises in the northeast and sets in the northwest, spending more than 12 hours in the sky. The winter Sun rises in the southeast and sets in the southwest, spending less than 12 hours in the sky. Both of these effects mean that northern latitudes receive more energy from the summer Sun, and summer days are warmer than winter days.



At summer solstice



At winter solstice

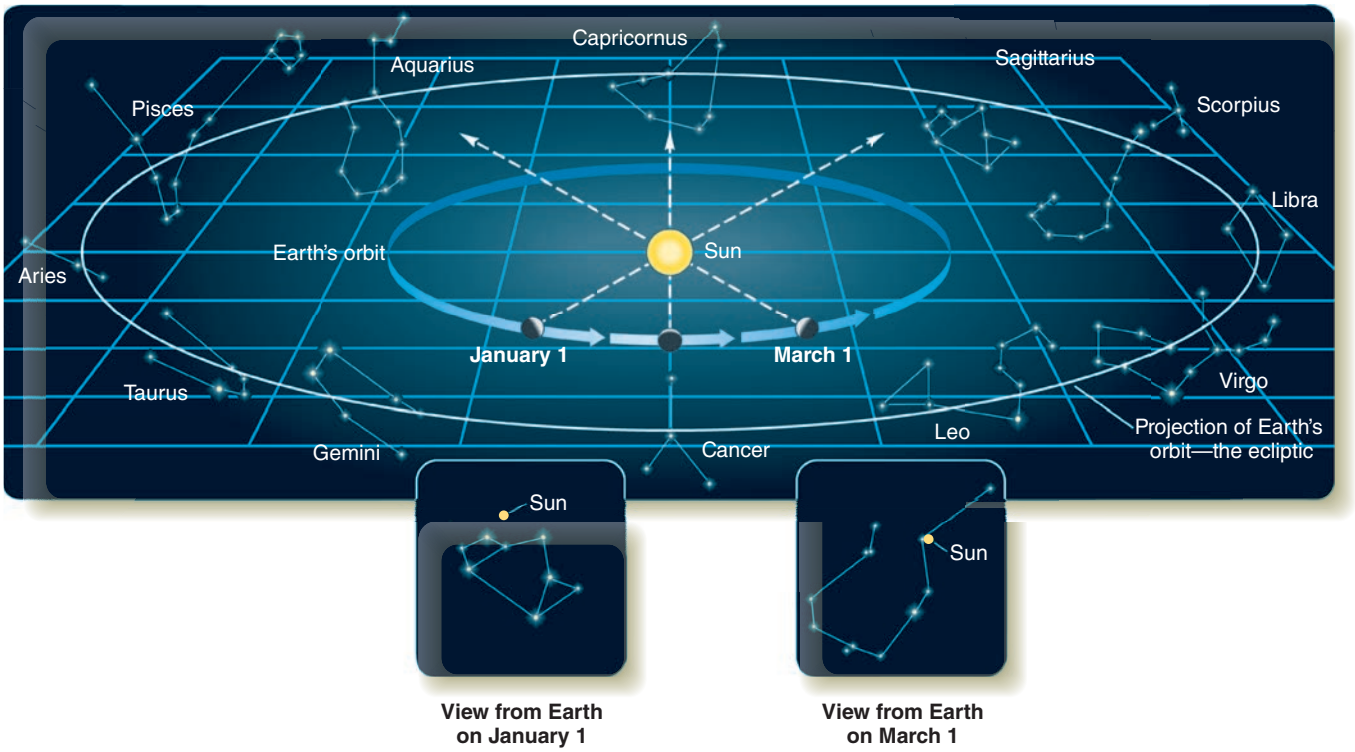


Earth at Northern Hemisphere winter solstice

1d On the day of the winter solstice in late December, Earth's Northern Hemisphere is inclined away from the Sun, and sunlight strikes the ground at an angle and spreads out. At southern latitudes, sunlight shines almost straight down and does not spread out. North America has cool weather and South America has warm weather.

Earth's orbit is only very slightly elliptical. About January 3, Earth is at **perihelion**, its closest point to the Sun, when it is only 1.7 percent closer than average. About July 5, Earth is at **aphelion**, its most distant point from the Sun, when it is only 1.7 percent farther than average. This small variation does not significantly affect the seasons.

NASA



▲ **Figure 2-9** Earth's orbit is a nearly perfect circle, but it is shown in an inclined view in this diagram and consequently looks oval. Earth's motion around the Sun makes the Sun appear to move against the background of the stars. Earth's orbit is thus projected on the sky as the path of the Sun, the ecliptic. If you could see the stars in the daytime, you would notice the Sun slowly crossing in front of the distant constellations as Earth moves along its orbit. (Sun symbols are not to scale.)

2 Both of Earth's hemispheres go through cycles of seasons because of changes in the amount of solar energy they receive at different times of the year. Circulation patterns in Earth's atmosphere keep the Northern and Southern Hemispheres mostly isolated from each other, and they exchange little heat. When one hemisphere receives more solar energy than the other, it grows rapidly warmer.

3 Notice that the seasons in Earth's Southern Hemisphere are reversed with respect to those in the Northern Hemisphere. Locations in the Southern Hemisphere experience winter from June 22 to September 22 and summer from December 21 to March 20.

Now you can set your friends straight if they mention two of the most **Common Misconceptions** about the seasons. First, the seasons are not caused by Earth moving closer to, or farther from, the Sun. If that were the cause, both of Earth's hemispheres would experience winter at the same time, when Earth is farthest from the Sun, and that's not what happens. Earth's orbit is nearly circular. Earth is actually closest to the Sun in January, but only 1.7 percent closer than the average, and 1.7 percent farther away than the average at its farthest, in July. That small variation isn't enough to cause noticeable seasons. Rather, the seasons arise because Earth's axis is not perpendicular to its orbit.

Here's a second **Common Misconception**: That it is easier to stand a raw egg on end on the day of the vernal equinox. Have you heard this one? Radio and TV personalities love to talk about it, but it just isn't true. It is one of the silliest misconceptions in science. You can stand a raw egg on end any day of the year if you have steady hands. (*Hint:* It helps to first shake the egg really hard to break the yoke inside so it can settle to the bottom.)

Throughout history, the cycle of the seasons, especially the dates of solstices and equinoxes, have been celebrated with rituals and festivals. Shakespeare's play *A Midsummer Night's Dream* describes the enchantment of the summer solstice night. (In many cultures, the equinoxes and solstices traditionally are taken to mark the midpoints, rather than the beginnings, of the seasons.) Many North American natives marked the summer solstice with ceremonies and dances. Early church officials placed Christmas Day in late December to coincide with a previous celebration of the winter solstice.

2-3c Motions of the Planets

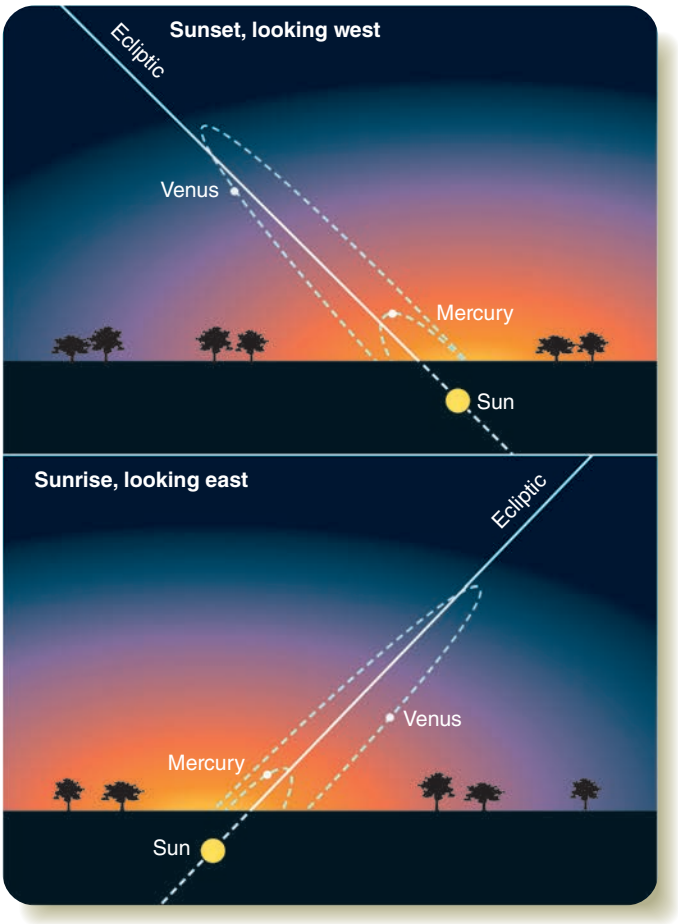
The planets of our Solar System produce no visible light of their own; they are visible only by reflected sunlight. Mercury, Venus, Mars, Jupiter, and Saturn are all easily visible to the unaided eye, but Uranus is usually too faint to be seen, and Neptune is never bright enough.

All of the planets of our Solar System, including Earth, move in nearly circular orbits around the Sun. If you were looking down on the Solar System from the north celestial pole, you would see the planets moving in the same counterclockwise direction around their orbits, with the planets farthest from the Sun moving the slowest. Seen from Earth, the outer planets move slowly eastward along the ecliptic. (You will learn about occasional exceptions to this eastward motion in Chapter 4.) In fact, the word *planet* comes from the Greek word meaning "wanderer." Mars moves completely around the ecliptic in slightly less than 2 years, but Saturn, being farther from the Sun, takes nearly 30 years.

Mercury and Venus also stay near the ecliptic, but they move differently from the other planets. They have orbits inside Earth's orbit, which means they are never seen far from the Sun in the sky. Observed from Earth, they move eastward away from the Sun and then back toward the Sun, crossing the near part of their orbit. They continue moving westward away from the Sun and then move back, crossing the far part of their orbit before they move to the east of the Sun again. To find one of these planets, you need to look above the western horizon just after sunset or above the eastern horizon just before sunrise. Venus is easier to locate because it is brighter and because its larger orbit carries it higher above the horizon than does Mercury's (**Figure 2-10**). Mercury's orbit is smaller and it can never be farther than 28 degrees from the Sun. Consequently, Mercury is hard to see against the glare of the sky near the Sun, and also it is often hidden in the clouds and haze near the horizon.

By tradition, any planet visible in the evening sky is called an **evening star**, even though planets are not stars. Similarly, any planet visible in the sky shortly before sunrise is called a **morning star**. Perhaps the most beautiful is Venus, which can become as bright as magnitude –4.7. As Venus moves around its orbit, it can dominate the western sky each evening for many weeks, but eventually its orbit carries it back toward the Sun, and it is lost in the haze near the horizon. In a few weeks, Venus reappears in the dawn sky, a brilliant morning star.

The cycles of the sky are so impressive that it is not surprising that people have strong feelings about them. Ancient peoples saw the motion of the Sun around the ecliptic as a powerful influence on their daily lives, and the motion of the planets along the ecliptic seemed similarly meaningful. The ancient superstition of astrology is based on the cycles of the Sun and planets around the sky. You have probably heard of the **zodiac**, a band around the sky extending about 9 degrees above and below the ecliptic in which the Sun, Moon, and planets are always found. The signs of the zodiac take their names from the 12 principal constellations along the ecliptic. A *horoscope* is just a diagram showing the location of the Sun, Moon, and planets around the ecliptic and their position above or below the horizon for a given date and time. Centuries ago, astrology was an important part of astronomy, but the two are now almost exact opposites—astronomy is a science that depends on evidence, and astrology is a superstition that survives despite evidence (**How Do We Know? 2-2**). The signs of the zodiac are no longer important in astronomy.



▲ **Figure 2-10** Mercury and Venus follow orbits that keep them near the Sun, and they are visible only soon after sunset or before sunrise when the brilliant Sun is hidden below the horizon. Venus takes 584 days to move from the morning sky to the evening sky and back again, but Mercury zips around in only 116 days.

2-4 Astronomical Influences on Earth's Climate

The seasons are produced by the annual motion of Earth around the Sun, but subtle changes in that motion can have dramatic effects on climate. You don't notice these changes during your lifetime, but, over thousands of years, they can bury continents under glaciers.

Earth has gone through ice ages when the worldwide climate was cooler and dryer and thick layers of ice covering the polar regions advanced multiple times partway to the equator and then retreated. The most recent ice age began about 3 million years ago

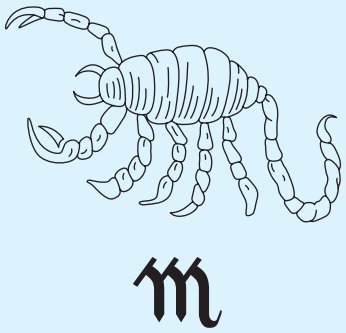
Pseudoscience

What is the difference between a science and a pseudoscience?

Astronomers have a low opinion of beliefs such as astrology, mostly because they are groundless but also because they *pretend* to be a science. They are pseudosciences, from the Greek *pseudo*, meaning false.

A **pseudoscience** is a set of beliefs that appears to include scientific ideas but fails to follow the most basic rules of science. For example, decades ago a claim (in other words, a hypothesis) was made that pyramidal shapes focus cosmic forces on anything underneath and might even have healing properties. Supposedly, a pyramid made of paper, plastic, or other materials would preserve fruit, sharpen razor blades, and do other miraculous things. Many books promoted the idea of the special power of pyramids, and this idea led to a popular fad.

A key characteristic of science is that its claims can be tested and verified. In this case, simple experiments showed that any shape, not just a pyramid, protects a piece of fruit from airborne spores and allows it to dry without rotting. Likewise, any shape prevents air flow and slows oxidation degradation of a razor blade's sharpness. Because experimental evidence contradicted the claim and because supporters of the hypothesis declined to abandon or revise their claims, you can recognize pyramid power as a pseudoscience. Disregard of contradictory evidence and alternate hypotheses is a sure sign of a pseudoscience.



Astrological sign and symbol for Scorpio; astrology may be the oldest pseudoscience.

Pseudoscientific claims can be self-fulfilling. For example, some believers in pyramid power slept under pyramidal tents to improve their rest. There is no physical mechanism by which such a tent could affect a sleeper, but because people wanted and expected the claim to be true they reported that they slept more soundly. Vague claims based on personal testimony that cannot be tested are another sign of a pseudoscience.

Astrology is probably the best-known pseudoscience. It has been tested over and over for centuries, and it simply does not work: It has been proven beyond any reasonable doubt that there is no connection between the positions of the Sun, Moon, and planets with people's personalities or events in their lives. Nevertheless, many people believe in astrology despite contradictory evidence.

Pseudosciences appeal to our need to understand and control the world around us. Some such claims involve medical cures, ranging from using magnetic bracelets and crystals to focus mystical power, to astonishingly expensive, illegal, and dangerous treatments for cancer. Even astrology can be harmful if you believe that your personality, choices, and possibilities are limited by your horoscope. Logic is a stranger to pseudoscience, but human fears and needs are not.

and is still going on—you know that because there is ice at the poles. You are living during one of the periodic episodes in the middle of an ice age when Earth grows slightly warmer and the glaciers melt back closer to the poles. The current warm period began about 12,000 years ago. Between ice ages, Earth is warmer and there are no ice sheets even at the poles. Scientists have found evidence of at least four ice ages in Earth's past. One occurred 2.5 billion years ago, but the other three that have been identified most clearly have all occurred in the last billion years. There were probably others, but evidence of early ice ages is usually erased by more recent ice sheets. The lengths of ice ages range from tens of millions to hundreds of millions of years.

During ice ages, the advance and retreat of glaciers has a complicated pattern that involves cycles of about 40,000 years and 100,000 years. These cycles have no connection with the current global warming that has produced changes in Earth's climate over just a few decades. Evidence shows that these slow cycles of the ice ages have an astronomical origin.

2-4a Milankovitch Climate Cycles: Hypothesis

Sometimes a hypothesis is proposed long before scientists can find the critical evidence to test it. That happened in 1920 when engineer and mathematician Milutin Milankovitch proposed what became known as the **Milankovitch hypothesis**, which states that small changes in the shape of Earth's orbit and axis inclination, along with a subtle effect of precession, could combine to influence Earth's climate and cause ice ages. You should examine each of these motions separately.

First, Earth's orbit is slightly eccentric. As you have learned, Earth's distance from the Sun varies by ±1.7 percent from its average during each year's orbit. You also know that the primary cause of seasons is the tilt of Earth's axis. The variation in distance from the Sun each year has some effect on the Sun's heating of Earth, but that is minor compared with the effect of the axial tilt. However, astronomers know that because of gravitational interactions with other planets, the eccentricity of Earth's orbit varies between 0 and about 3 percent over a period of

approximately 100,000 years. When the orbit has low eccentricity (is almost perfectly circular), summers in some locations can be not warm enough to melt all of the snow and ice from the previous winter, tending to make glaciers grow larger.

A second factor is the inclination of Earth's equator to its orbit, currently 23.4 degrees. Because of gravitational tugs of the Moon, Sun, and planets, this angle varies between 22 and almost 25 degrees with a period of 41,000 years. When the inclination is less, the summer/winter contrast is less and glaciers tend to grow.

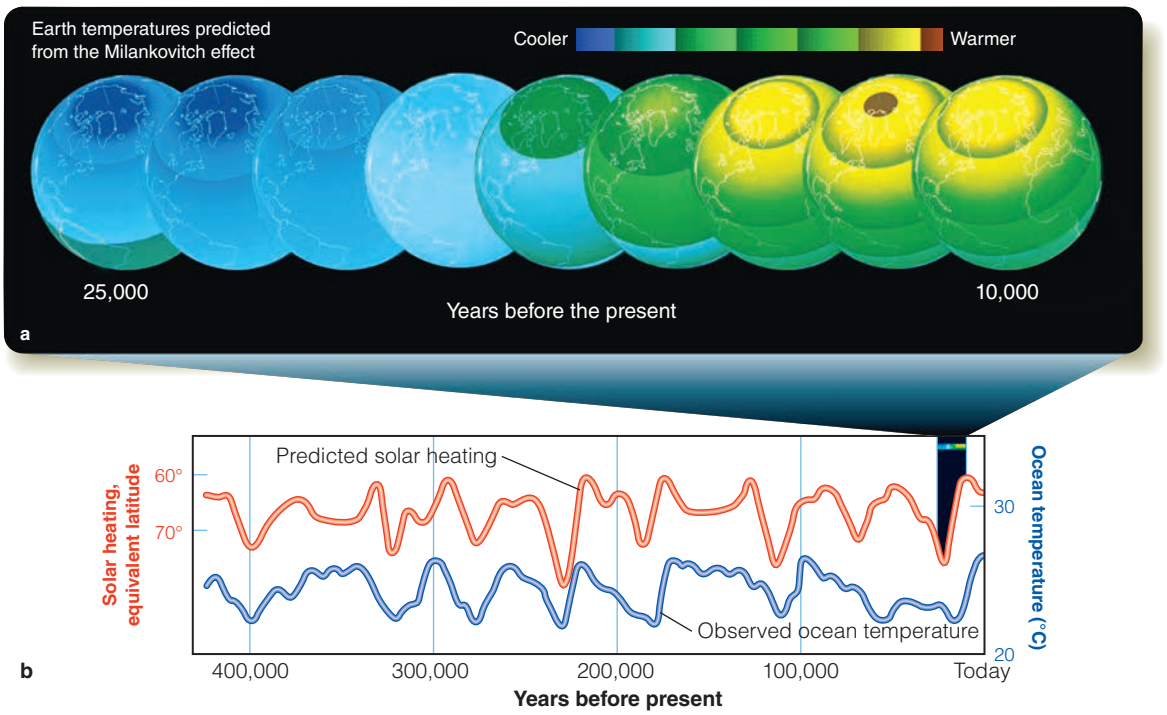
A third factor is also involved. As you learned in the previous section, precession causes Earth's axis to sweep around a circle on the celestial sphere with a period of roughly 26,000 years. As a result, the points in Earth's orbit where a given hemisphere experiences each season gradually change. Northern Hemisphere summers now occur when Earth is farther from the Sun than average, making northern summers slightly cooler. (In this context, the Northern Hemisphere is the most important part of the globe: Most of the landmass where ice can accumulate is located in the north.) In 13,000 years, half a precession cycle from now, northern summers will occur on the other side of Earth's orbit where Earth is closer to the Sun. Northern summers will be slightly warmer, more able to melt all of the previous winter's snow and ice and prevent the growth of glaciers.

In 1920, Milankovitch proposed that these three cyclical factors combine with each other to produce complex periodic variations in Earth's climate that result in the advance and retreat of glaciers (**Figure 2-11a**). However, little evidence was available at the time to test the hypothesis, and scientists treated it with deep skepticism.

2-4b Milankovitch Climate Cycles: Evidence

By the mid-1970s, Earth scientists were able to collect the data that Milankovitch had lacked. Oceanographers drilled deep into the seafloor to collect long cores of sediment. In the laboratory, geologists could take samples from different depths in the cores and determine the age of the samples and the temperature of the oceans when they were deposited on the seafloor. From this, scientists constructed a history of ocean temperatures that convincingly matched the predictions of the Milankovitch hypothesis (**Figure 2-11b**).

The evidence seemed very strong, and by the 1980s the Milankovitch hypothesis was widely considered the leading hypothesis to explain climate cycles during the current ice age. But science follows a (mostly unstated) set of rules that holds that a hypothesis must be tested repeatedly against all available



▲**Figure 2-11** (a) Calculations based on the Milankovitch hypothesis can be used to predict temperatures on Earth over time. The warming illustrated by the Earth globes shown here took place from 25,000 to 10,000 years ago and ended the last glacial advance. (b) Over the last 400,000 years, changes in ocean temperatures measured from fossils found in sediment layers on the seafloor approximately match calculated changes in solar heating. The globes in panel (a) illustrate events in only a short segment near the recent (right) end of the timelines.

Scientific Arguments

How is a scientific argument different from a legal argument?

People in the legal profession and in science have a different meaning for the word *argument* than the sense of "verbal battle" that is often used in casual conversation. Lawyers and scientists use *argument* to mean a summary of evidence and principles leading to a conclusion.

A prosecuting attorney constructs an argument to persuade the judge or jury that the accused is guilty; a defense attorney in the same trial constructs an argument to persuade the same judge or jury toward the opposite conclusion. Neither prosecutor nor defender is obliged to consider anything that weakens their respective cases, and, in the United States, the defendant has a constitutional right not to say anything that might incriminate him or her. Lawyers rarely, if ever, include a statement in a closing argument about the possibility that that they might be wrong.

Scientists construct arguments because they want to test their own ideas and give an accurate explanation of some aspect of nature. For example, in the 1960s, biologist E. O. Wilson presented a scientific argument to show that ants communicate by smells. The argument included a description of his careful observations and the ingenious experiments he had conducted to test his hypothesis. And, this is truly important: Wilson also considered other evidence and alternate explanations for ant communication. Scientists can include any evidence or hypothesis that supports their claim, but they must observe one fundamental rule of professional science: They must be as honest as possible; they must include all of the known evidence and all of the hypotheses previously proposed. Unlike lawyers, scientists must explicitly account for the possibility that they might be wrong.

Scientists publish their work in the form of scientific arguments, but they also think in scientific arguments. If, in thinking through his argument, Wilson had found a contradiction, he would have known he was on the wrong track. That is why scientific arguments must be complete and honest. Scientists who ignore inconvenient evidence or brush aside other explanations are only fooling themselves.

A good scientific argument gives you all the information you need to decide for yourself whether the argument is correct. Wilson's study of ant communication is now widely understood and is being applied to other fields such as pest control and telecommunications networks.

Eye of Science/Science Source

Through a series of scientific arguments, biologists have learned that ants communicate with a large vocabulary of smells.

evidence. In 1988, scientists discovered some apparently contradictory evidence.

For 500,000 years rainwater has collected in a deep crack in a cave in Nevada called Devils Hole. That water has deposited the mineral calcite in layer on layer on the walls of the crack. It isn't easy to get to, and scientists had to dive with scuba gear to drill out samples of the calcite, but it was worth the effort. Back in the laboratory, they could determine the age of each layer in their core samples and the temperature of the rainwater that had formed the calcite in each layer. That gave them a history of temperatures at Devils Hole that spanned many thousands of years, and the results were a surprise. The new data seemed to show that Earth had begun warming up thousands of years too early for the last glacial retreat to have been caused by the Milankovitch cycles.

These contradictory findings were confusing because we humans naturally prefer certainty, but such circumstances are common in science. The disagreement between the ocean floor samples and the Devils Hole samples triggered a scramble to understand the problem. Were the age determinations of one or the other set of samples wrong? Were the calculations of prehistoric temperatures wrong? Or were scientists misunderstanding the significance of the evidence?

In 1997, a new study of the ages of the samples confirmed that those from the ocean floor were correctly dated. But the same study found that the ages of the Devils Hole samples were also correct. Evidently the temperatures at Devils Hole record local climate changes in the region that is now the southwestern United States. The ocean floor samples record global climate changes, and they fit well with the Milankovitch hypothesis. This has given scientists renewed confidence in the Milankovitch hypothesis in a general sense, but also made them aware that some regions of Earth do not exactly follow trends seen for the rest of the planet. Although it is widely accepted today, the Milaknovitch hypothesis is still being tested whenever scientists can find more evidence.

As you review this section, notice that it is a **scientific argument**, a careful presentation of hypothesis and evidence in a logical discussion. **How Do We Know? 2-3** expands on the ways scientists organize their ideas in logical arguments. Also,

throughout this book, many chapter sections end with a short feature titled "Practicing Science." These are case studies featuring one or two review questions that can be answered by imagining yourself as a scientist analyzing measurements, inventing hypotheses, constructing scientific arguments, and so on. You can use the "Practicing Science" features to review chapter material but also to practice thinking like a scientist.

Practicing **SCIENCE**

Why was it critical, in testing the Milankovitch hypothesis about ice age climate change mechanisms, for scientists to determine the ages of ocean sediment? Ocean floors accumulate sediment in thin layers year after year. Scientists can drill into the ocean floor and collect cores of those sediment layers, and from chemical tests they can find the temperature of the seawater when each layer was deposited. Those determinations of past temperatures can be used as reality checks in the scientific argument regarding the Milankovitch hypothesis, but only if the ages of the sediment layers are determined correctly. When a conflict arose with evidence from Devils Hole in Nevada, the age determinations of the both the ocean floor and Devils Hole samples were carefully reexamined and found to be correct. After reviewing all of the evidence, scientists concluded that the ocean core samples did indeed support the Milankovitch hypothesis.

WHAT ARE WE?

Along for the Ride

Humans have spread over the surface of Earth like a thin coat of paint. If you use your astronomical perspective, you can see that we humans are confined to the surface of our world.

The rotation of Earth creates a cycle of day and night that controls everything from TV schedules to the chemical workings of our brains. We wake and sleep within that 24-hour cycle of light and dark. Furthermore, Earth's orbital motion around the Sun, combined with the inclination of its axis, creates a yearly cycle of seasons, and we humans, along with every other living thing on Earth, have evolved to thrive within those extremes of temperature. We protect ourselves from the largest extremes and inhabit most of Earth's land area, hunting, gathering, and growing crops within the cycle of the seasons.

In recent times, we have begun to understand that conditions on Earth's surface are not entirely stable. Slow changes in Earth's motions and orientation of the planet produce irregular cycles of glaciation. All of recorded history, including the creation of all of the cities, roads, and communication networks that paint our globe, has occurred since the last glacial retreat ended only 12,000 years ago, so we humans have no recorded experience of Earth's coldest climate. We have never known the icy side of our planet's personality. We are along for the ride and enjoying Earth's good times.

Study and Review

SUMMARY

2-1 Stars and Constellations

- Although the **constellations** currently used by astronomers originated in Middle Eastern and Greek mythology, the names are Latin. Even modern constellations, added to fill in the spaces between the ancient figures, have Latin names. Named groups of stars such as the Big Dipper or Orion's Belt that are not complete constellations are called **asterisms**. Astronomers now divide the sky into 88 constellations, defined in 1928 by the **International Astronomical Union (IAU)**.
- The names of individual stars usually come from old Arabic, though modern astronomers often refer to a bright star by its constellation plus a Greek letter assigned according to its brightness within the constellation.
- Astronomers describe the brightness of stars using the **magnitude scale**. First-magnitude stars are brighter than second-magnitude stars, which are brighter than third-magnitude stars, and so on. The magnitude describing what you see when you look at a star in the sky is its **apparent visual magnitude (m_V)**, which includes only types of light visible to the human eye and also does not take into account the star's distance from Earth.
- **Flux** is a measure of light energy striking one square meter per second. The magnitude of a star is related directly to the flux of light received on Earth from that star.

$$\frac{F_A}{F_B} = (2.51)^{(m_B - m_A)} \qquad \text{(Eq. 2-1a)}$$

$$m_B - m_A = 2.5 \log\left(\frac{F_A}{F_B}\right) \qquad \text{(Eq. 2-1b)}$$

2-2 The Sky and Celestial Motions

- The **celestial sphere** is a **scientific model** of the sky, to which the stars appear to be attached. Because Earth rotates eastward, the celestial sphere appears to rotate westward on its axis.
- The **north** and **south celestial poles** are the pivots on which the sky appears to rotate, and they define the four cardinal directions around the **horizon**: the **north**, **south**, **east**, and **west points**. The point directly overhead is the **zenith**, and the point on the sky directly underfoot is the **nadir**. The **celestial equator**, which is an imaginary line around the sky above Earth's equator, divides the sky into Northern and Southern Hemisphere.
- As the celestial sphere is curved, the distances between the stars "on" the sky are angular, not linear, distances. These **angular distances**, measured in degrees, **arc minutes**, and **arc seconds**, are not directly related to the true distance between the objects measured in units such as kilometers (km) or light-years (ly). The angular distance across an object is its **angular diameter**.
- What you see of the celestial sphere depends on your latitude. **Circumpolar constellations** are those close enough to a celestial pole that they do not appear to rise from the east and set in the west. The angular distance from the horizon to the north celestial pole as measured from the north point always equals your latitude.
- **Precession** is caused by the gravitational forces of the Moon and Sun acting on the equatorial bulge of the spinning Earth, causing Earth's axis to sweep around in a conical motion like the motion of a wobbling top's axis. Earth's axis precesses with a period of 26,000 years. As a result, the positions of the celestial poles and celestial equator move slowly against the background of the stars.

2-3 Sun and Planets

- The **rotation** of Earth on its axis produces the daily cycle of day and night, and the **revolution** of Earth around the Sun produces the annual cycle of the seasons.
- Because Earth orbits the Sun, the Sun appears to move eastward along the **ecliptic**, through the constellations, completing a circuit of the sky in a year. Because the ecliptic is tipped 23.4 degrees to the celestial equator, the Sun spends half the year north of the celestial equator and half the year south of the celestial equator.
- The seasons are caused by the angle of sunlight and how long the Sun is in the sky at different times of year. The beginning of spring, summer, winter, and fall are marked by the **vernal equinox**, the **summer solstice**, the **autumnal equinox**, and the **winter solstice**. In its orbit around the Sun, Earth is slightly closer to the Sun at **perihelion** in January and slightly farther away from the Sun at **aphelion** in July. This change in distance to the Sun has almost no effect on Earth's seasons.
- The planets appear to move generally eastward along the ecliptic. Planets visible in the sky at sunset are traditionally called **evening stars**, and planets visible in the dawn sky are called **morning stars** even though they are not actually stars. The locations of the Sun and planets along the **zodiac** are the basis for the ancient **pseudoscience** (or false science), known as astrology.

2-4 Astronomical Influences on Earth's Climate

- According to the **Milankovitch hypothesis**, slow changes in the shape of Earth's orbit, the angle of axis tilt, and axis orientation can alter the planet's heat balance and cause the cycle of ice advances and retreats during an ice age. Evidence found in seafloor samples and other locations support the hypothesis, and the hypothesis is widely accepted today.
- Scientists routinely test their own ideas by organizing theory and evidence into a **scientific argument**.

REVIEW QUESTIONS

1. Why are most of the constellations that were invented in modern times composed of faint stars or located in the southern sky?
2. How does the Greek letter designation of a star give you clues both to its location and its apparent brightness?
3. Which is the asterism and which is the constellation: Orion and Orion's belt? Name another asterism/constellation combination.
4. From your knowledge of star names and constellations, which of the following stars in each pair is probably brighter? Explain your answers.
   a. Alpha Ursae Majoris; Epsilon Ursae Majoris
   b. Alpha Telescopii; Alpha Orionis
5. How did the magnitude system originate in a classification of stars by apparent brightness?
6. What does the word *apparent* mean in *apparent visual magnitude?*
7. What does the word *visual* mean in *apparent visual magnitude?*
8. Do the apparent visual magnitude numbers of two stars tell us which one is farther from Earth? Explain your answer.
9. Do the apparent visual magnitude numbers of two stars tell us which one is bigger? Explain your answer.
10. Do the apparent visual magnitude numbers of two stars tell us which star shines more light on the same area of Earth in the same time interval? Explain your answer.
11. Why doesn't a magnitude difference of one mean that the corresponding flux ratio is also one?
12. If Star B has $m_B = -2$ and Star A $m_A = 0$, from which star does Earth receive more flux?
13. In what ways is the celestial sphere a scientific model?
14. Is the precessing top shown in Figure 2-7a an example of a scientific model? If so, which parts of the model are true and which parts are not necessarily true?
15. How would you define the celestial poles and equator if Earth did not rotate but kept one hemisphere constantly facing the Sun during its yearly orbit?
16. Where would you need to go on Earth to see the celestial equator very near your horizon?
17. Where would you go on Earth if you wanted to be able to see both the north celestial pole and the south celestial pole at the same time?
18. Your zenith is at your east point and your nadir is at your west point. Are you sitting, squatting, standing, lying facedown on the ground, or lying faceup on the ground? If your arms are outstretched and perpendicular to your body, in which direction(s) are your arms pointing?
19. Why does the number of circumpolar constellations depend on the latitude of the observer?
20. Explain two reasons winter days are colder than summer days.
21. How does the date of the beginning of summer in Earth's Southern Hemisphere differ from the date in the Northern Hemisphere?
22. If it is the first day of spring in your hemisphere, what day is it in the opposite hemisphere?
23. It is the first day of summer. Will the days start getting longer tomorrow, start getting shorter tomorrow, or will the Sun stay up in the sky tomorrow for as long as it did today?
24. How much flux from the Sun does the Northern Hemisphere of Earth receive compared to the Southern Hemisphere on the first day of fall at noon?
25. Why does the eccentric shape of Earth's orbit make winter in Earth's Northern Hemisphere different from winter in Earth's Southern Hemisphere?

26. **How Do We Know?** How can a scientific model be useful if it is *not* a true description of nature?
27. **How Do We Know?** Why is astrology a pseudoscience?
28. **How Do We Know?** Why must a scientific argument dealing with some aspect of nature take all of the known evidence into account?

ACTIVE INQUIRY QUESTIONS

1. Using stars from one or more of the "official" constellations, create an asterism that is significant to the culture of your school.
2. You discover an ancient document that lists the location of buried treasure as 36° 24′ 11.9″ N by 25° 25′ 35.4″ E. You look on a map to see approximately where that is, but when you start on your trip you find that your GPS is broken. How can you most easily determine your latitude without GPS? Much harder question: How can you determine your longitude without GPS?
3. You are packing for a vacation on a planet orbiting another star that is much like the Sun. Why might you want some information about the planet's orbit size and axis tilt to know what to pack?

PROBLEMS

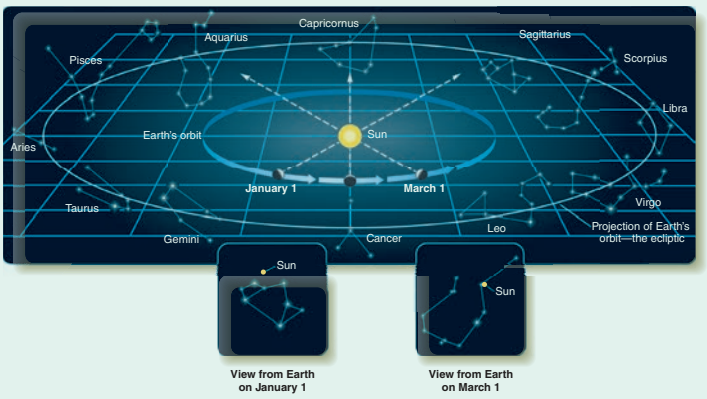
1. Star A has a magnitude of 9.5; Star B, 5.5; and Star C, 2.5. Which is brightest? Which are visible to the unaided eye? Which pair of stars has a flux ratio of about 16?
2. If one star is 7.3 times brighter than another star, how many magnitudes brighter is it?
3. If light from one star is 251 times brighter (has 251 times more flux) than light from another star, what is their difference in magnitudes?
4. If Earth receives twice as much light per unit area per unit time from Star A compared to Star B, what is the apparent visual magnitude difference between the stars? Which star is apparently brighter, Star A or Star B?
5. If Earth receives one-third as much light per unit area per unit time from Star A compared to Star B, what is the apparent visual magnitude difference between the stars? Which star is apparently brighter, Star A or Star B?
6. If in the previous problem Star A has an apparent visual magnitude of 5, what is the apparent visual magnitude of Star B? Which can you see on a clear night in your sky using your eyes: Star A, Star B, Star A and Star B, or neither Star A nor Star B?
7. If two stars differ by 8 magnitudes, what is their flux ratio?
8. If two stars differ by 5.6 magnitudes, what is their flux ratio?
9. If Star A is magnitude 1.0 and Star B is magnitude 9.6, which is brighter and by what factor?
10. By what factor is the full moon brighter than Venus at its brightest? (*Hint*: Refer to Figure 2-6.)
11. What is the angular distance from the north celestial pole to the point on the sky called the vernal equinox? To the summer solstice?
12. If you are at latitude 40 degrees north of Earth's equator, what is the angular distance from the northern horizon up to the north celestial pole? From the southern horizon down to the south celestial pole?
13. If you are at latitude 30 degrees north of Earth's equator, what is the angular distance from your zenith to the north celestial pole? From your nadir to the north celestial pole?
14. How many precession periods are in one cycle of Earth's axis inclination variation? In one cycle of Earth's orbit eccentricity variation? In the time span shown in Figure 2-11b, how many periods or fractions of periods did the Earth's axis precess, nod, and Earth's orbit change shape? Of the three periods, which is likely to have the most effect on the changes shown in Figure 2-11?

SENSE OF PROPORTION

1. Arrange the following in order of increasing apparent brightness: Sirius; a full moon; Polaris; Venus; the Sun; the faintest star you can see without a telescope
2. Arrange the following in order of increasing length of time: one precession cycle; one rotation of Earth; one revolution of Earth around the Sun; one revolution of the Moon around Earth; any one of Earth's four seasons

LEARNING TO LOOK

1. Find the Big Dipper in the star trails photograph on the left side of **The Sky Around You**.
2. Look at **The Sky Around You**, item 1a. In the looking south illustration, is Canis Major a circumpolar constellation? Why or why not?
3. Look at the view from Earth on March 1 in Figure 2-9, shown here. Is the view from Earth's nighttime side or daytime side? How do you know? Which asterism or constellation is shown in this image?
4. Look at Figure 2-9, shown here. If you see Sagittarius high in your night sky on June 20 and today is your birthday, what is your zodiac constellation?



3 Moon Phases and Eclipses

NASA/Rami Daud

▲ A sequence of images spanning about 2 hours shows a lunar eclipse visible from the United States in September 2015. From left to right, the full moon is increasingly covered by Earth's shadow. In the image at far right, the Moon is shown fully shadowed, illuminated only by the orange glow of sunlight refracted through Earth's atmosphere.

In the previous chapter, you learned about daily and yearly cycles of the Sun's appearance in Earth's sky. Now you can focus on phenomena of the next brightest object in the sky, the Moon. The Moon moves against the background of stars, changing its appearance in a monthly cycle and occasionally producing spectacular events called *eclipses*. This chapter will help you answer four important questions about Earth's natural satellite:

- Why does the Moon go through phases?
- What causes a lunar eclipse?
- What causes a solar eclipse?
- How can eclipses be predicted?

Understanding the phases of the Moon and eclipses will exercise your scientific imagination as well as help you enjoy the sight of the Moon crossing the sky.

Once you have an understanding of the sky as it appears from Earth, you will be ready to read the next chapter about how Renaissance astronomers analyzed motions of the Sun, Moon, and planets; used their imaginations; and came to a revolutionary conclusion—that Earth is a planet.

> **O, swear not by the moon, the fickle moon, the inconstant moon, that monthly changes in her circle orb, Lest that thy love prove likewise variable.**
>
> WILLIAM SHAKESPEARE, *ROMEO & JULIET*

SINCE ANCIENT TIMES, superstitious people have associated the Moon with insanity. The word *lunatic* comes from a time when even doctors thought that the insane were "moon-struck." Of course, the Moon does not cause madness, but it is bright and beautiful, moves visibly across the constellations, and is associated with eclipses and tides, so people might *expect* it to have a dramatic effect on them.

3-1 The Changeable Moon

Starting this evening, begin looking for the Moon in the sky. You might have to wait for almost a month before you find it appearing at a convenient time, but then, as you watch for the Moon on successive evenings, you will see it following its orbit around Earth and cycling through its phases as it has done for billions of years (**Figure 3-1**).

3-1a The Moon's Orbital Motion

Just as Earth would be seen to revolve counterclockwise around the Sun if viewed from the direction of the north celestial pole, the Moon revolves counterclockwise around Earth. Because the Moon's orbit is tipped about 5 degrees from the plane of Earth's orbit, the Moon's path is always near the ecliptic, sometimes slightly north of it and sometimes slightly south of it.

The Moon moves rapidly against the background of the constellations. If you watch the Moon for just an hour, you can see it move eastward by slightly more than its angular diameter. The Moon is about 0.5 degree in angular diameter, and it moves eastward a bit more than 0.5 degree per hour. In 24 hours, it moves 13 degrees. Thus, each night you see the Moon about 13 degrees eastward of its location the night before.

As the Moon orbits around Earth, its appearance changes from night to night in a monthly cycle.

3-1b The Cycle of Moon Phases

The changing appearance of the Moon as it revolves around Earth, called the **lunar phase** cycle, is one of the most easily observed phenomena in astronomy. Study **The Phases of the Moon** and notice three important points and two new terms:

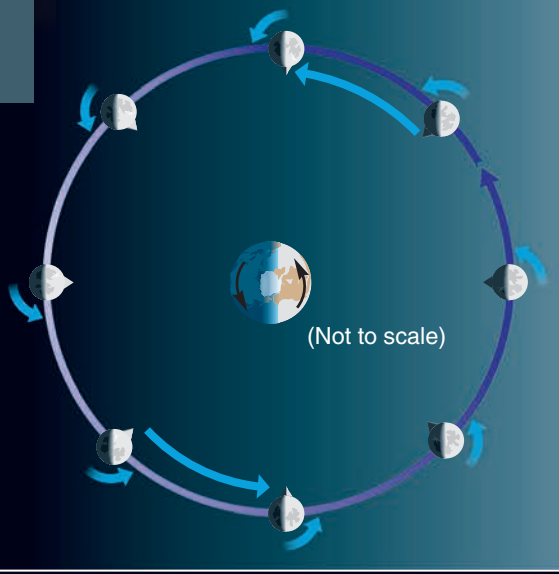
1. The Moon always keeps the same side facing Earth. "The Man in the Moon" is produced by the familiar features on the Moon's near side, but you never see the far side of the Moon from Earth.
2. The changing shape of the Moon as it passes through its cycle of phases is produced by sunlight illuminating different portions of the side of the Moon you can see.
3. Notice the difference between the orbital period of the Moon around Earth relative to the background stars (the *sidereal period,* pronounced "si-DARE-ee-al," referring to stars) versus the length of the lunar phase cycle (the *synodic period*). That difference is a good illustration of how your view from Earth is produced by the combined motions of Earth and other celestial bodies such as the Sun and Moon.

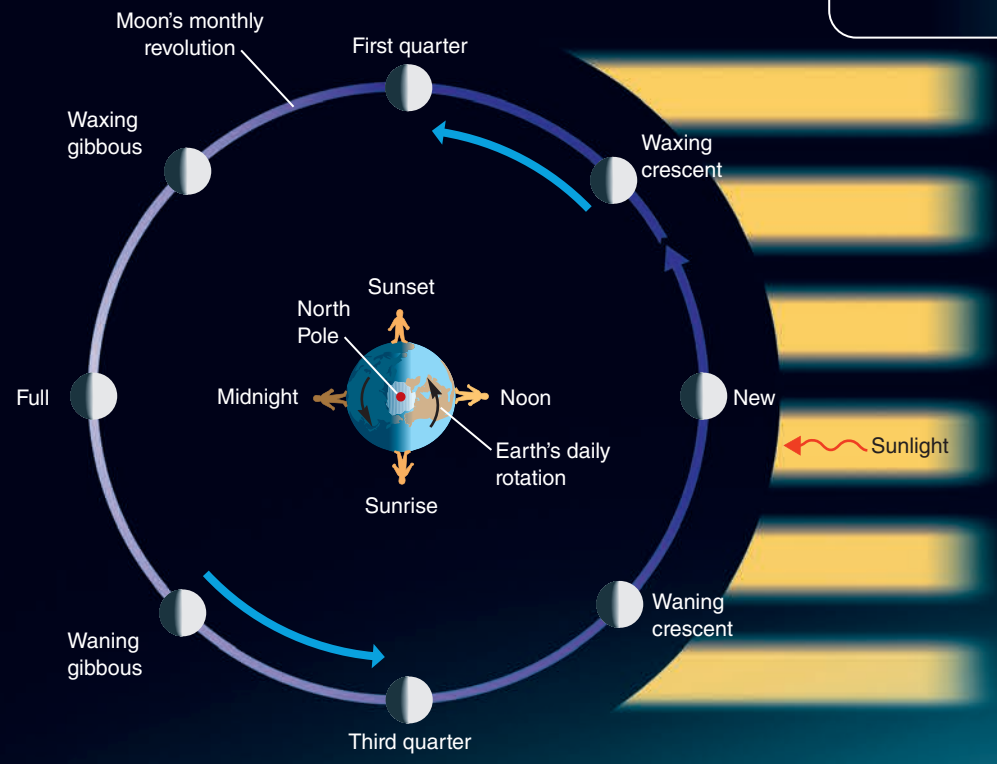
© UC Regents / Lick Observatory

▲ **Figure 3-1** In this sequence of lunar phase snapshots taken at intervals of 1 day, the Moon cycles through its phases from crescent to full to crescent. From Earth you see the same face of the Moon, the same mountains, craters, and plains, at all times, but the changing direction of sunlight changes what part is illuminated and produces the lunar phases.

The Phases of the Moon

1 As the Moon orbits Earth, it rotates so that the same side always faces Earth, as shown at right. Consequently, you always see the same features on the Moon from Earth, and you never see the far side of the Moon. A mountain on the Moon that points at Earth will always point at Earth as the Moon rotates on its axis and revolves around Earth.

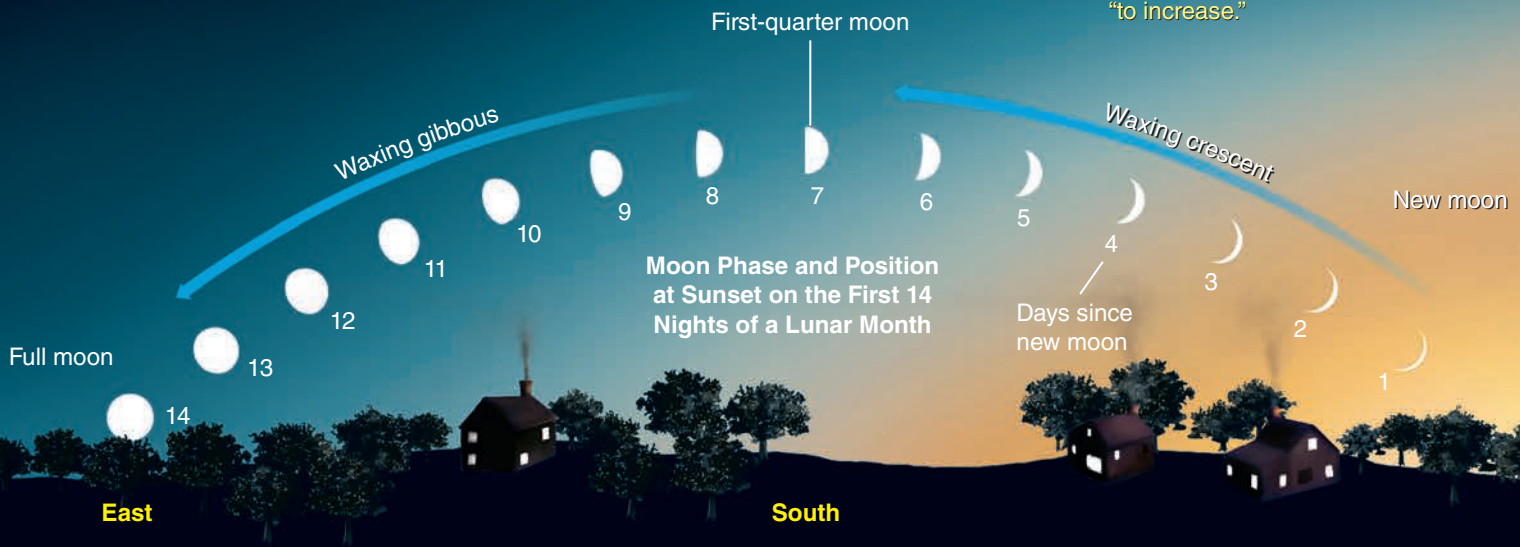




2 As seen at left, sunlight always illuminates half of the Moon. Because you see different amounts of this sunlit side, you see the Moon go through a cycle of phases. At the phase called *new moon*, when the Moon is between Earth and the Sun, sunlight illuminates the far side of the Moon, and the side you see is in darkness. In fact, at new moon you cannot see the Moon at all in contrast to the bright daytime sky near the Sun. At full moon, when Earth is between the Moon and the Sun, the side of the Moon you can see from Earth is fully lit, and the far side is in darkness. How much of the Moon you see illuminated depends on where the Moon is in its monthly orbit around Earth.

The observer's time of day depends on his or her location on Earth relative to the Sun.

2a The first two weeks of the Moon's monthly cycle are shown below by its position and phase as seen at sunset on 14 successive evenings. As the illuminated part of the Moon grows larger from new to full, it is said to *wax*, an old-fashioned word meaning "to increase."

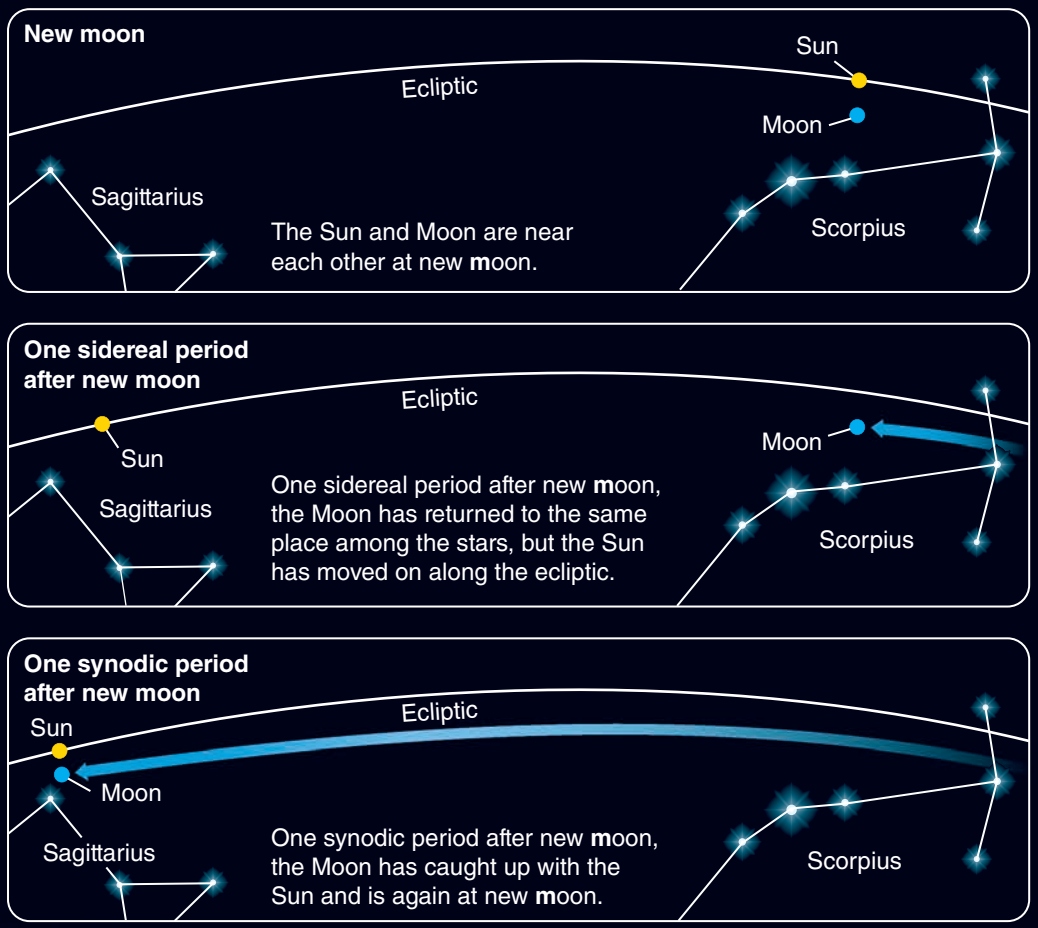


Moon Phase and Position at Sunset on the First 14 Nights of a Lunar Month

3 The Moon orbits eastward around Earth in 27.3 days, the Moon's **sidereal period**. This is how long the Moon takes to circle the sky once and return to the same position relative to the stars.

A complete cycle of lunar phases takes 29.5 days, the Moon's **synodic period**. (Synodic comes from the Greek words for "together" and "path.") To see why the synodic period is longer than the sidereal period, study the star charts at the right.

Although you think of the lunar cycle as being about 4 weeks long, it is actually 1.53 days longer than 4 weeks. The fact that our calendar divides the year into months (literally "moonths") about 30 days long originates from the 29.5-day synodic cycle of the Moon.

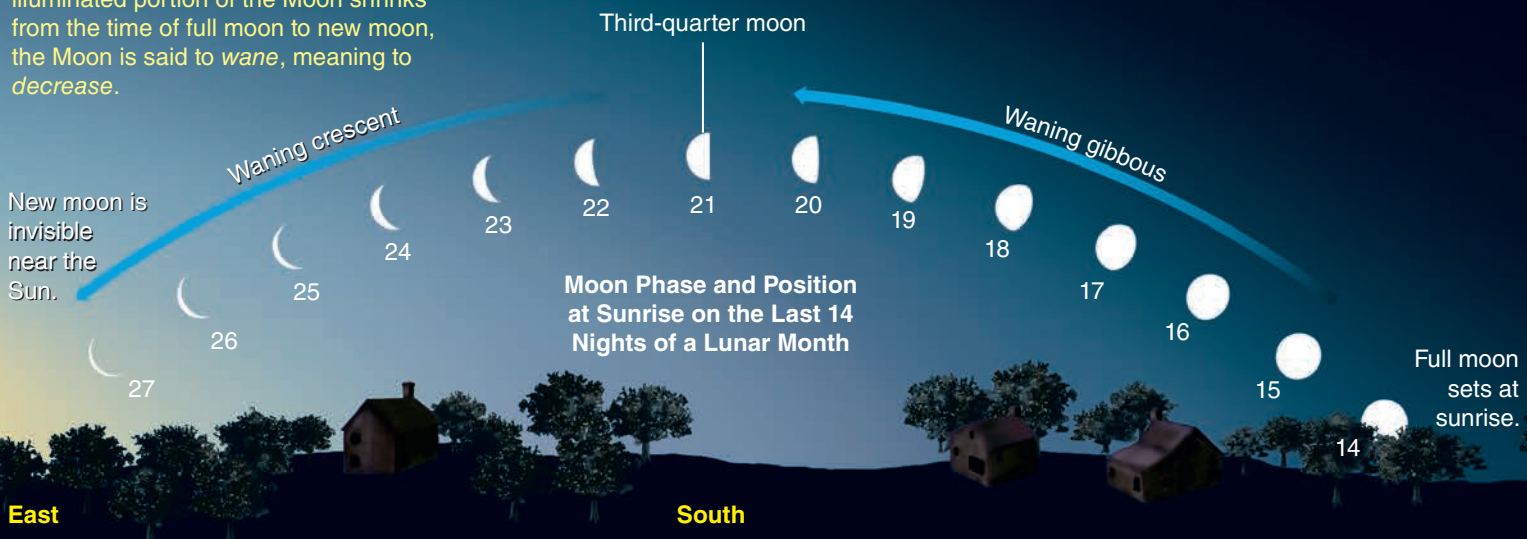

New moon
Ecliptic
Sun
Moon
Sagittarius
Scorpius
The Sun and Moon are near each other at new moon.
One sidereal period after new moon
Ecliptic
Sun
Moon
Sagittarius
Scorpius
One sidereal period after new moon, the Moon has returned to the same place among the stars, but the Sun has moved on along the ecliptic.
One synodic period after new moon
Ecliptic
Sun
Moon
Sagittarius
Scorpius
One synodic period after new moon, the Moon has caught up with the Sun and is again at new moon.

You can use the diagram on the opposite page to determine when the Moon rises and sets at different phases.

TIMES OF MOONRISE AND MOONSET

Phase	Moonrise	Moonset
New	Dawn	Sunset
First quarter	Noon	Midnight
Full	Sunset	Dawn
Third quarter	Midnight	Noon

2b The last 2 weeks of the Moon's monthly cycle are shown below by its position and phase as seen at sunrise on 14 successive mornings. As the illuminated portion of the Moon shrinks from the time of full moon to new moon, the Moon is said to *wane*, meaning to *decrease*.



Moon Phase and Position at Sunrise on the Last 14 Nights of a Lunar Month

The phases of the Moon are dramatic, and they have attracted lots of peculiar ideas. You have probably heard a number of **Common Misconceptions** about the Moon. Sometimes people are surprised to see the Moon in the daytime sky, and they think something has gone wrong. No, the gibbous (just past full) moon is often visible in the daytime. Quarter moons and especially crescent moons can also be in the daytime sky but are harder to see when the Sun is above the horizon and the sky is bright. You may also hear people mention "the dark side of the moon," but you can assure them that this is a misconception; there is no permanently dark side. Any location on the Moon is sunlit for two weeks and is in darkness for two weeks as the Moon revolves and rotates. Finally, you have probably heard one of the strangest misconceptions about the Moon: that people tend to act up at full moon. Careful statistical studies of records from schools, prisons, hospitals, and so on show that it isn't true. There are always a few people who misbehave, and the Moon has nothing to do with it.

For billions of years, "the Man in the Moon" has looked down on Earth. People in the first civilizations saw the same monthly cycle of phases that you see, and even the dinosaurs may have noticed the changing phases of the Moon. Occasionally, however, the Moon displays more complicated moods during a lunar eclipse.

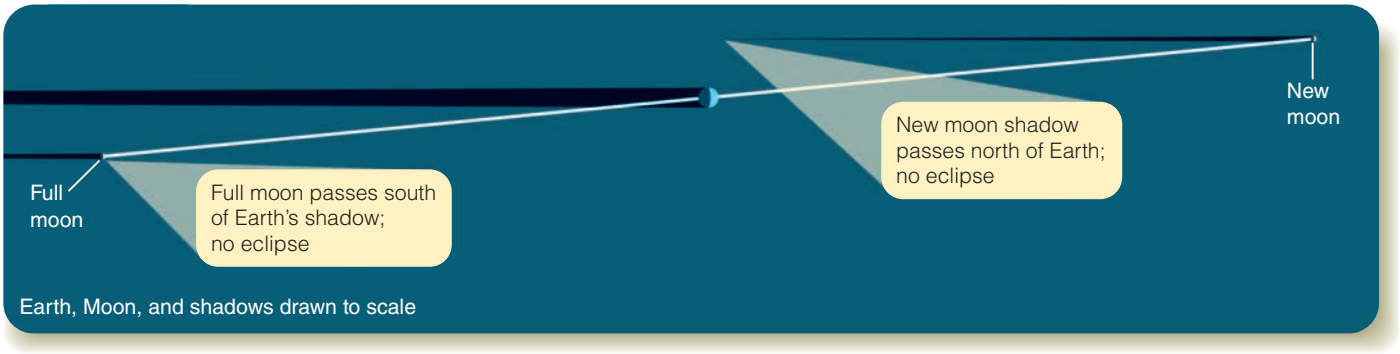
3-2 Lunar Eclipses

In cultures all around the world, the sky is a symbol of order and power, and the Moon is a regular counter of the passing days and months. It is not surprising that people are startled and sometimes worried when, once in a while, they see the full moon become dark and coppery-red colored. Such events are neither mysterious nor frightening once you understand how they arise. To begin, you can think about Earth's shadow.

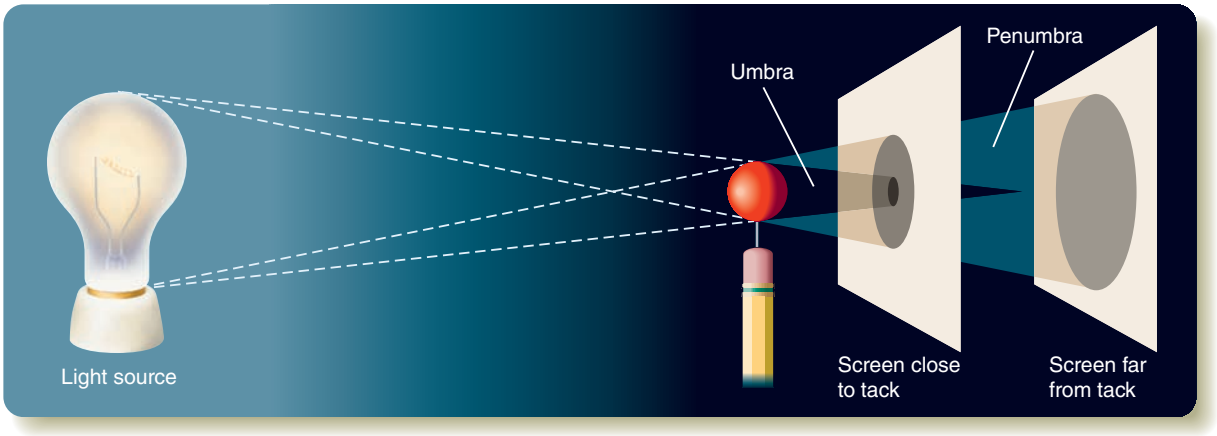
3-2a Earth's Shadow

As you just learned, the orbit of the Moon is tipped only a few degrees from the plane of Earth's orbit around the Sun. Earth's shadow points directly away from the Sun in the plane of Earth's orbit. A **lunar eclipse** can occur at full moon (and only full moon) if the Moon's path carries it through the shadow of Earth, sunlight is blocked, and the Moon grows dim temporarily. This is somewhat unusual because most full moons pass north or south of ("above" or "below") Earth's shadow, and there is no eclipse (**Figure 3-2**). The conditions that allow eclipses to occur will be explained in detail later in this chapter.

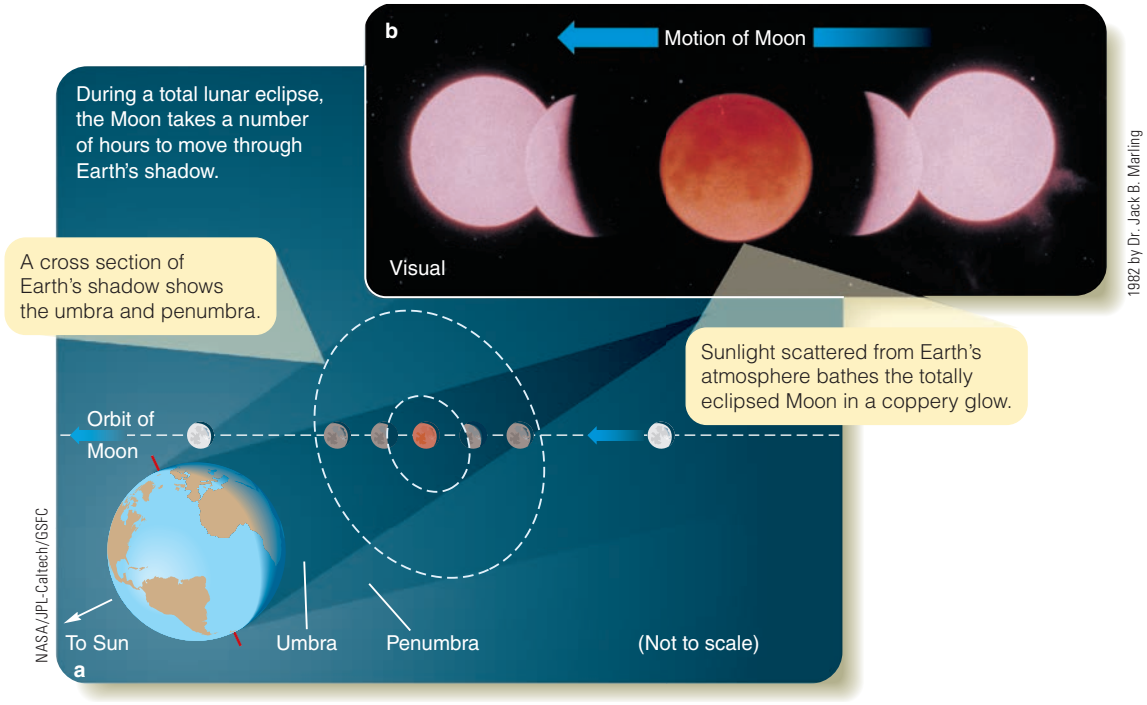
A shadow consists of two parts (**Figure 3-3**). The **umbra** is the region of total shadow. If you were floating in a spacesuit in


Full moon
Full moon passes south of Earth's shadow; no eclipse
New moon shadow passes north of Earth; no eclipse
New moon
Earth, Moon, and shadows drawn to scale

▲ **Figure 3-2** The Moon's orbit around Earth is tilted relative to Earth's orbit around the Sun, so the long, thin shadows of Earth and Moon usually miss each other. In most months there are no eclipses.



▲ **Figure 3-3** The shadow cast by a map tack can be used to understand the shadows of Earth and the Moon. The umbra is the region of total shadow; the penumbra is the region of partial shadow.


b
Motion of Moon
Visual
1982 by Dr. Jack B. Marling
During a total lunar eclipse, the Moon takes a number of hours to move through Earth's shadow.
A cross section of Earth's shadow shows the umbra and penumbra.
Sunlight scattered from Earth's atmosphere bathes the totally eclipsed Moon in a coppery glow.
Orbit of Moon
NASA/JPL-Caltech/GSFC
To Sun
Umbra
Penumbra
(Not to scale)
a

▲ **Figure 3-4** (a) In this diagram of a total lunar eclipse, the Moon passes from right to left through Earth's shadow. (b) A multiple-exposure photograph shows the Moon passing through the umbra of Earth's shadow. A longer exposure was used to record the Moon while it was totally eclipsed. The Moon's path appears curved in the photo because of photographic effects.

the umbra of Earth's shadow, the Sun would be completely hidden behind Earth, and you would not be able to see any part of the Sun's bright disk. If you drifted into the **penumbra**, however, you would see part of the Sun peeking around the edge of Earth, so you would be in partial shadow. In the penumbra, the Sun is partly but not completely blocked.

The umbra of Earth's shadow is more than three times longer than the distance to the Moon and points directly away from the Sun. A giant screen placed in the shadow at the average distance of the Moon would reveal a dark umbral shadow about 2.5 times the diameter of the Moon. The faint outer edges of the penumbra would mark a circle about 4.6 times the diameter of the Moon. Consequently, when the Moon's orbit carries it through Earth's shadow, the shadow is plenty large enough for the Moon to become completely immersed.

3-2b Total Lunar Eclipses

Once or twice a year, the Moon's orbit carries it through the umbra of Earth's shadow, and if for some interval the Moon is completely within the umbra, you see a **total lunar eclipse** (Figure 3-4a). As you watch the eclipse begin, the Moon first moves into the penumbra and dims slightly; the deeper it moves into the penumbra, the more it dims. Eventually, the Moon reaches the umbra, and you see the umbral shadow darken part, then all, of the Moon.

When the Moon is totally eclipsed, it does not disappear completely. While it is in the umbra it receives no direct sunlight, but the Moon is illuminated by some sunlight that is refracted (bent) through Earth's atmosphere. If you were on the Moon during **totality**, you would not see any part of the Sun because it would be entirely hidden behind Earth. However, you would see Earth's atmosphere lit from behind by the Sun. The red glow from this ring of sunsets and sunrises around the circumference of Earth shines into the umbra of Earth's shadow, making the umbra not completely dark. That glow illuminates the Moon during totality and makes it seem reddish in color, as shown in Figure 3-4b and in Figure 3-5.

How dim the totally eclipsed Moon becomes depends on a number of things. Total lunar eclipses tend to be darkest when the Moon's orbit carries it through the center of the umbra. Also, if Earth's atmosphere is especially cloudy in the regions that are bending light into the umbra, the Moon will be darker than during an average eclipse. An unusual amount of dust in Earth's atmosphere (for example, from volcanic eruptions) can also cause an exceptionally dark or especially red eclipse.

The exact timing of a lunar eclipse depends on where the Moon crosses Earth's shadow. If it crosses through the center of the umbra, the eclipse will have maximum length. For such an eclipse, the Moon spends about an hour crossing the penumbra and then another hour entering the darker umbra. Totality can

Celestron International

▲ **Figure 3-5** During a total lunar eclipse, the Moon turns a coppery-red color. In this photo, the Moon is darkest toward the lower right, the direction toward the center of the umbra. The edge of the Moon at upper left is brighter because it is near the edge of the umbra.

last as long as 1 hour 45 minutes, followed by the emergence of the Moon into the penumbra, plus another hour as it emerges into full sunlight. Thus, a total lunar eclipse can take nearly 6 hours from start to finish.

3-2c Partial and Penumbral Lunar Eclipses

Because the Moon's orbit is inclined by a bit more than 5 degrees to the plane of Earth's orbit around the Sun, the Moon does not always pass through the center of the umbra (look back to Figure 3-2). If the Moon passes a bit too far north or south, it may only partially enter the umbra, and we see a **partial lunar eclipse**. The part of the Moon that remains in the penumbra receives some direct sunlight, and the glare is usually great enough to prevent us from seeing the faint red glow of the part of the Moon in the umbra. For that reason, partial lunar eclipses are not as beautiful as total eclipses.

If the orbit of the Moon carries it far enough north or south of the umbra, the Moon may pass through only the penumbra and never reach the umbra. Such **penumbral lunar eclipses** are not dramatic at all. In the partial shadow of the penumbra, the Moon is only partially dimmed. Most people glancing at a penumbral eclipse would not notice any difference from a full moon.

Total, partial, or penumbral lunar eclipses are interesting events in the night sky and are not difficult to observe. When the full moon passes through Earth's shadow, the eclipse is visible from anywhere on Earth's dark side. Consult **Table 3-1** to find the next lunar eclipse visible in your part of the world.

Table 3-1 Total and Partial Eclipses of the Moon, 2018 through 2028[a]

Date	Time of Mid-Eclipse (UTC)[b]	Length of Totality (Hr:Min)	Length of Eclipse[c] (Hr:Min)
2018 January 31	13:31	1:16	3:23
2018 July 27	20:23	1:43	3:55
2019 January 21	05:13	1:02	3:17
2019 July 16	21:32	Partial	2:58
2021 May 26	11:20	0:15	3:07
2021 November 19	09:04	Partial	3:28
2022 May 16	04:13	1:25	3:27
2022 November 8	11:00	1:25	3:40
2023 October 28	20:15	Partial	1:17
2024 September 18	02:45	Partial	1:03
2025 March 14	7:00	1:05	3:38
2025 September 7	18:13	1:22	3:29
2026 March 3	11:35	0:58	3:27
2026 August 28	04:14	Partial	3:18
2028 December 31	16:53	1:11	3:29

[a]There will be no total or partial lunar eclipses during 2020 and 2027.
[b]Times are Universal Time. Subtract 5 hours for Eastern Standard Time, 6 hours for Central Standard Time, 7 hours for Mountain Standard Time, and 8 hours for Pacific Standard Time. For Daylight Saving Time (mid-March through early November), add 1 hour to Standard Time. Lunar eclipses that occur between sunset and sunrise in your time zone will be visible, and those at midnight will be best placed.
[c]Does not include penumbral phase.
Source: NASA Goddard Space Flight Center

Practicing SCIENCE

What would a total lunar eclipse look like if Earth had no atmosphere? As a way to test and improve their understanding, scientists often experiment with their ideas by imagining changing one part of a system and trying to figure out what would happen as a result. This is sometimes called a "thought experiment." In this example, the absence of an atmosphere around Earth would mean that no sunlight would be bent toward the eclipsed Moon, and it would not glow red. The Moon would be very dark in the sky during totality.

Now try a new thought experiment; imagine changing a different part of the Earth–Moon–Sun system and guess the result. ***What would a lunar eclipse look like if the Moon and Earth were the same diameter?***

3-3 Solar Eclipses

For millennia, cultures worldwide have understood that the Sun is the source of life, so you can imagine the panic people felt at the terrible sight of the Sun gradually disappearing in the middle of the day. Many imagined that the Sun was being devoured by a monster (**Figure 3-6**). Modern scientists must use their imaginations to visualize how nature works, but with a key difference: They test their ideas against reality (**How Do We Know? 3-1**).

Scientific Imagination

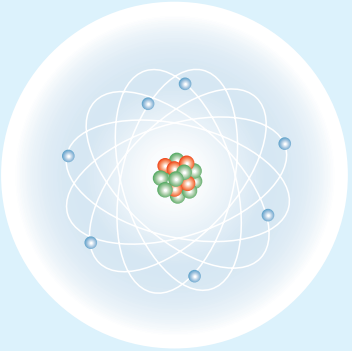
How do scientists produce hypotheses to test?

Good scientists are invariably creative people with strong imaginations who can study raw data about some invisible aspect of nature such as an atom and construct mental pictures as diverse as a plum pudding or a planetary system. These scientists share the same human impulse to understand nature that drove ancient cultures to imagine eclipses as serpents devouring the Sun.

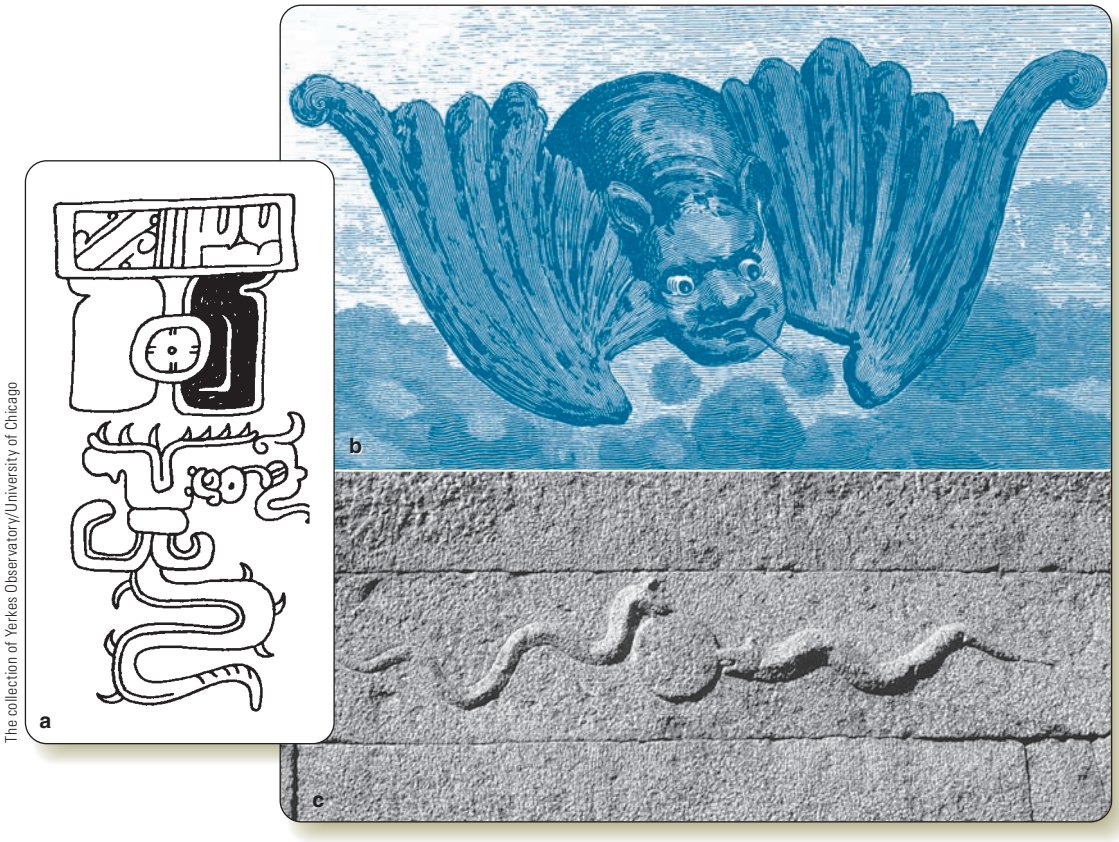
As the 20th century began, physicists were busy trying to imagine what an atom was like. No one can see an atom, but English physicist J. J. Thomson used what he knew from his experiments and his imagination to create an image of what an atom might be like. He suggested that an atom was a ball of positively charged material with negatively charged electrons distributed throughout like plums in a plum pudding.

The key difference between using a plum pudding to represent the atom and a hungry serpent to represent an eclipse is that the plum pudding model was based on experimental data and could be tested against new evidence. As it turned out, Thomson's student, Ernest Rutherford, performed ingenious new experiments and showed that atoms can't be made like plum puddings. Rather, his data led him to imagine an atom as a tiny positively charged nucleus surrounded by negatively charged electrons, much like a tiny version of the Solar System with planets circling the Sun. Later experiments confirmed that Rutherford's description of atoms is closer to reality, and it has become a universally recognized symbol for atomic energy.

Ancient cultures pictured the Sun being devoured by a serpent. Thomson, Rutherford, and scientists like them used their scientific imaginations to visualize natural processes and then test and refine their ideas with new experiments and observations. The critical difference is that scientific imagination is continuously tested against reality and is revised when necessary.



A model image of the atom as electrons orbiting a small nucleus has become the symbol for atomic energy.



The collection of Yerkes Observatory/University of Chicago

T. Scott Smith

◀ **Figure 3-6** (a) A 12th-century Mayan symbol believed to represent a solar eclipse. The black-and-white Sun symbol hangs from a rectangular sky symbol, and a voracious serpent approaches from below. (b) The Chinese representation of a solar eclipse shows a monster, usually described as a dragon, flying in front of the Sun. (c) This wall carving from the ruins of a temple in Vijayanagara in southern India symbolizes a solar eclipse as two snakes approach the disk of the Sun.

Courtesy of Allan Meyer

▲ **Figure 3-7** A total solar eclipse occurs when the Moon crosses in front of the Sun and hides its brilliant surface, allowing a glimpse of the Sun's atmosphere. This image was taken on August 21, 2017 from near Madras, Oregon.

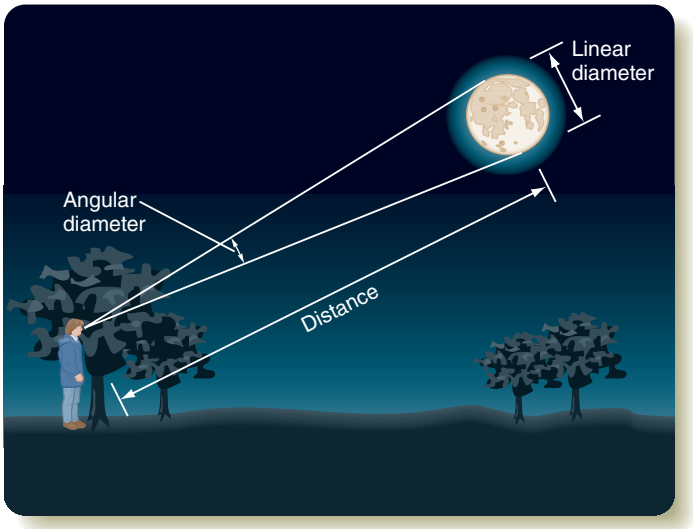
A **solar eclipse** occurs when the Moon moves between Earth and the Sun. If the Moon covers the disk of the Sun completely, you see a spectacular **total solar eclipse** (Figure 3-7). If, from your location, the Moon covers only part of the Sun, you see a less dramatic **partial solar eclipse**. During a solar eclipse people in one place on Earth may see a total eclipse while people only a few hundred kilometers away see a partial eclipse.

The geometry of a solar eclipse is quite different from that of a lunar eclipse. You can begin by considering how big the Sun and Moon look in the sky.

3-3a The Angular Diameters of the Sun and Moon

Solar eclipses are spectacular because Earth's Moon happens to have nearly the same angular diameter as the Sun, so it can cover the Sun's disk almost exactly. You learned about angular diameter in the Chapter 2 Concept Art: The Sky Around You; now you can consider how to determine the Moon's angular diameter from its size and distance.

Linear diameter is simply the distance between an object's opposite sides. You use linear diameter when you order a 16-inch pizza—the pizza is 16 inches across. In contrast, the angular diameter of an object is the angle formed by lines extending toward you from opposite edges of the object and meeting at your eye (Figure 3-8). Clearly, the farther away an object is, the smaller its angular diameter.



▲ **Figure 3-8** The angular diameter of an object is related to both its linear diameter and its distance.

To find the angular diameter of the Moon, you need to use the **small-angle formula**. That formula expresses the relationship of the linear (true) diameter, the angular (apparent) diameter, and the distance, of any object, whether it is a pizza or the Moon. If you know two of those quantities, you can find the third one by cross-multiplying. This formula is used very often in astronomy, and you will encounter it many times in later chapters:

KEY EQUATION 3-1 Small-Angle Formula

$$\frac{\text{angular diameter (in arc seconds)}}{2.06 \times 10^5} = \frac{\text{linear diameter}}{\text{distance}}$$

In the small-angle formula, you must always use the same units for distance and linear diameter. This version of the small-angle formula uses arc seconds as the unit of angular diameter. (The constant 2.06×10^5 in the formula is the number of arc seconds in one radian.)

Example 1: The Moon's linear diameter is 3480 km (2160 mi), and its average distance from Earth is 384,000 km (both values have been rounded to a precision of three digits). What is its angular diameter? ***Solution:*** Because the Moon's linear diameter and distance are both given in the same units, kilometers, you can put them directly into the small-angle formula:

$$\frac{\text{angular diameter}}{2.06 \times 10^5} = \frac{3480 \text{ km}}{384{,}000 \text{ km}}$$

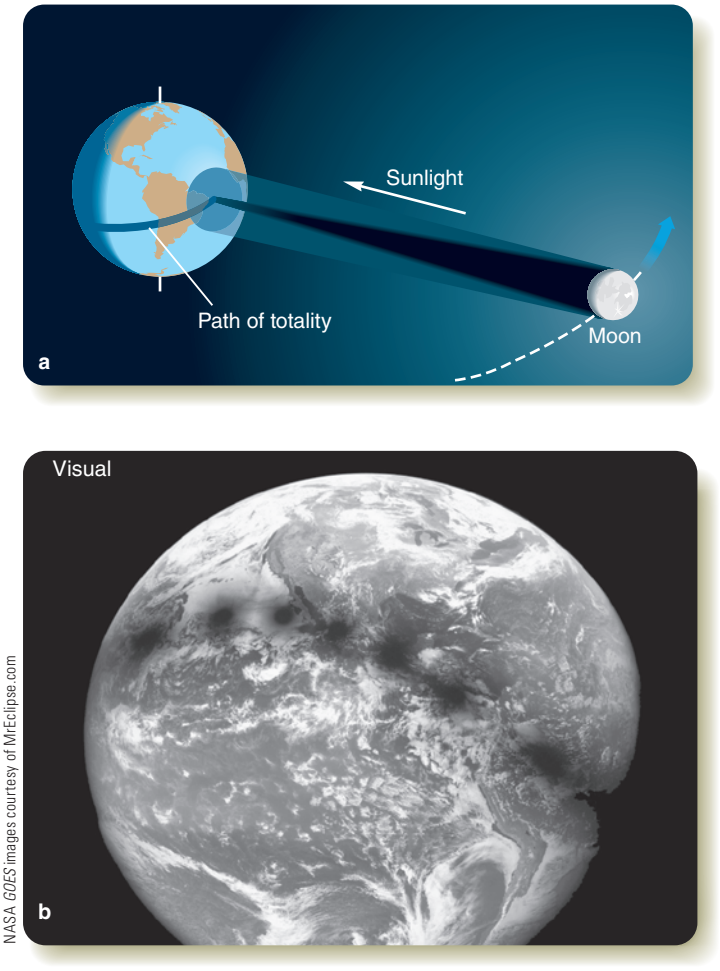
When you do the calculation you will find that the angular diameter of the Moon is 1870 arc seconds (rounded to three digits of precision). If you divide by 60, you get 31 arc minutes; dividing by 60 again, you get about 0.5 degree. The Moon's orbit is slightly elliptical, so the Moon can sometimes look a bit larger or smaller, but its angular diameter is always close to 0.5 degree. It is a **Common Misconception** that the Moon is larger when it is on the horizon. Certainly the rising full moon looks big when you see it on the horizon, but that is an optical illusion. In reality, the Moon is the same size on the horizon as when it is high overhead.

Now, do another small-angle calculation. ***Example 2:*** What is the angular diameter of the Sun? ***Solution:*** The Sun has a linear diameter of 1.39×10^6 km and its average distance from Earth is 1.50×10^8 km. If you put these numbers into the small-angle formula, you will find that the Sun has an angular diameter of 1910 arc seconds, which is about 32 arc minutes, or 0.5 degree.

By fantastic good luck, you live on a planet with a moon that is almost exactly the same angular diameter as its sun. Thanks to that coincidence, when the Moon passes in front of the Sun, it is almost exactly the right size to cover the Sun's brilliant surface but leave the Sun's atmosphere visible.

3-3b The Moon's Shadow

To see a solar eclipse, you have to be in the Moon's shadow. Like Earth's shadow, the Moon's shadow consists of a central umbra of total shadow and a penumbra of partial shadow. The Moon's umbral shadow produces a spot of darkness no more than 270 km (170 mi) in diameter on Earth's surface. (The exact size of the umbral shadow depends on the location of the Moon in its elliptical orbit and the angle at which the shadow strikes Earth.) The combination of Earth's rotation with the Moon's orbital motion causes the shadow to rush across Earth at speeds of at least 1800 km/h (1120 mph), sweeping out a **path of totality** (Figure 3-9). People lucky enough to be in the path of totality will see a total eclipse of the Sun while the umbral spot sweeps over them. Observers



NASA *GOES* images courtesy of MrEclipse.com

▲ **Figure 3-9** (a) The umbra of the Moon's shadow sweeps from west to east across Earth, and observers in the path of totality see a total solar eclipse. Those outside the umbra but inside the penumbra see a partial eclipse. (b) Eight photos made by a weather satellite have been combined to show the Moon's shadow moving across the eastern Pacific, Mexico, Central America, and Brazil during an eclipse in 1991.

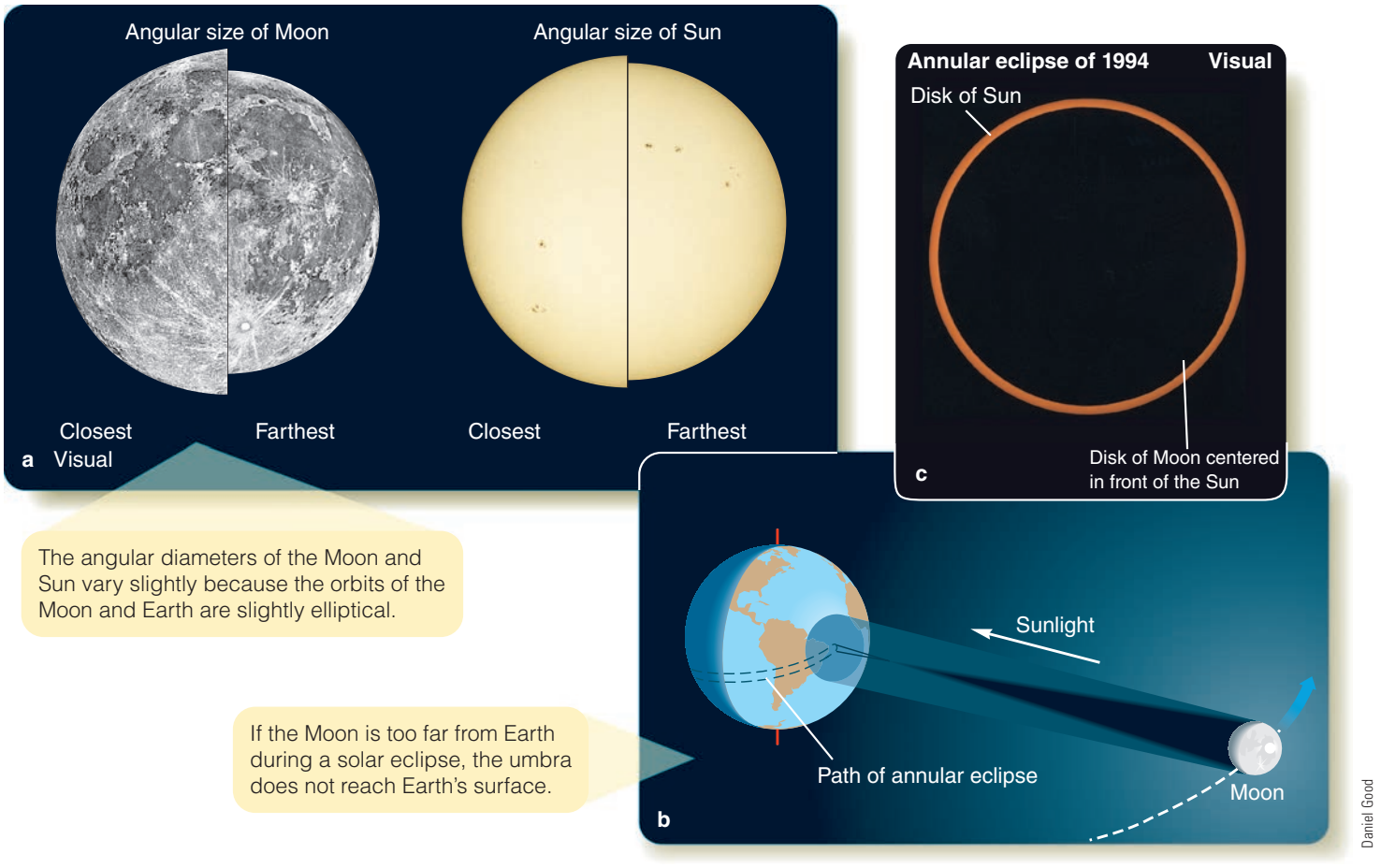
just outside the path of totality will see a partial solar eclipse as the penumbral shadow sweeps over their location. Those living even farther from the path of totality will see no eclipse.

The shapes of the Moon's orbit around Earth and of Earth's orbit around the Sun affect the details of what Earth-bound observers see during eclipses. The orbit of the Moon is slightly elliptical, and its distance from Earth varies. When it is at **apogee**, its farthest point from Earth, the Moon's angular diameter is 5.5 percent smaller than average, and when it is at **perigee**, its closest point to Earth, its angular diameter is 5.5 percent larger than average. Another factor is Earth's slightly elliptical orbit around the Sun. When Earth is closest to the Sun in January, the Sun looks 1.7 percent larger in angular diameter; and when Earth is farthest from the Sun in July, the Sun looks 1.7 percent smaller. As a result of those effects, sometimes the disk of the Moon is not big enough to cover the Sun as seen from Earth's surface. You can see in Figure 3-10 that this is equivalent to saying that sometimes the Moon's umbral shadow is not long enough to reach Earth. If the Moon crosses in front of the Sun when the Moon's disk is smaller in angular diameter than the Sun's, it produces an **annular eclipse**, a solar eclipse in which a ring (or annulus) of light is visible around the disk of the Moon.

Total solar eclipses are rare unless you are willing to leave home to see one. If you stay in one location, you will see a total solar eclipse on average about once every 360 years. On the other hand, some people are eclipse chasers: They plan years in advance and travel halfway around the world to place themselves in the path of totality. Table 3-2 shows the date and location of solar eclipses over the next few years.

3-3c Features of Solar Eclipses

A solar eclipse begins when you first see the edge of the Moon encroaching on the Sun. This is the moment when the edge of the penumbra sweeps over your location.


Angular size of Moon
Angular size of Sun
Closest
Farthest
Closest
Farthest
a Visual
The angular diameters of the Moon and Sun vary slightly because the orbits of the Moon and Earth are slightly elliptical.
Annular eclipse of 1994
Visual
Disk of Sun
Disk of Moon centered in front of the Sun
c
Sunlight
If the Moon is too far from Earth during a solar eclipse, the umbra does not reach Earth's surface.
Path of annular eclipse
Moon
b
Daniel Good

▲ **Figure 3-10** (a) Because the angular diameter of the Moon and the Sun vary slightly, the disk of the Moon is sometimes too small to cover the disk of the Sun. (b) That means the umbra of the Moon does not reach Earth, and the eclipse is annular, meaning a ring ("annulus") of the Sun's disk can be seen around the Moon. (c) In this photograph of an annular eclipse in 1994, the dark disk of the Moon is almost exactly centered on the bright disk of the Sun.

Table 3-2 Total and Annular Eclipses of the Sun, 2019 through 2028[a]

Date	Total/Annular (T/A)	Time of Mid-Eclipse (UTC)[b]	Maximum Length of Total or Annular Phase (Min:Sec)	Area of Visibility
2019 July 2	T	19:24	4:33	Pacific, South America
2019 December 26	A	05:19	3:39	Southeast Asia, Pacific
2020 June 21	A	06:41	00:38	Africa, South Asia, East Asia
2020 December 14	T	16:15	02:10	South Pacific, South America, South Atlantic
2021 June 10	A	10:43	3:51	North America, Arctic
2021 December 4	T	07:35	1:54	Antarctica, South Atlantic
2023 April 20	A/T[c]	04:18	1:16	Southeast Asia, Philippines, Indonesia, Australia
2023 October 14	A	18:00	5:17	United States, Central America, South America
2024 April 8	T	18:18	4:28	North America, Central America
2024 October 2	A	18:46	7:25	Pacific, South America
2026 February 17	A	12:13	2:20	South America, Atlantic, South Africa, Antarctica
2026 August 12	T	17:47	2:18	North America, North Atlantic, Arctic, Europe, Africa
2027 February 6	A	16:01	7:51	South America, Antarctica, South Atlantic, South Africa
2027 August 2	T	10:08	6:23	Africa, Europe, West and South Asia
2028 January 26	A	15:09	10:27	South America, North America, Atlantic, Europe
2028 July 22	T	02:57	5:10	Southeast Asia, Australia and New Zealand, Pacific

[a]There will be no total or partial solar eclipses in 2018 and 2025.
[b]Times are Universal Time. Subtract 5 hours for Eastern Standard Time, 6 hours for Central Standard Time, 7 hours for Mountain Standard Time, and 8 hours for Pacific Standard Time. For Daylight Saving Time (mid-March through early November), add 1 hour to Standard Time.
[c]Hybrid eclipse: begins as annular, becomes total, ends as annular.
Source: NASA Goddard Space Flight Center

During the partial phases of a solar eclipse, the Moon gradually covers the bright disk of the Sun (Figure 3-11). Totality begins as the last sliver of the Sun's bright surface disappears behind the Moon. This is the moment when the edge of the umbra sweeps over your location. So long as any of the Sun is visible, the landscape remains bright, but, as the last of the Sun disappears, darkness falls in a few seconds. Automatic streetlights come on, drivers switch on their headlights, and birds go to roost. The darkness of totality depends on a number of factors, including the weather at the observing site, but it is usually dark enough to make reading difficult.

The totally eclipsed Sun is a spectacular sight. With the Moon covering the bright surface of the Sun, called the **photosphere**, you can see the Sun's faint outer atmosphere, called the **corona**, glowing with a pale white light faint enough that you can safely look at it directly. The corona is made of hot, low-density gas that is given a wispy appearance by the solar magnetic field, as shown in the bottom frame of Figure 3-11 and even more so in Figure 3-7. Also visible just above the photosphere is a thin layer of bright gas called the **chromosphere**. The chromosphere is often marked by eruptions on the solar surface called **prominences** (Figure 3-12a) that glow with a clear pink color because of the high temperature of the gases involved. A large prominence can be wider than three times the diameter of Earth. (The nature of the photosphere, chromosphere, corona, and prominences as components of the Sun's atmosphere will be described in detail in Chapter 8.)

Totality during a solar eclipse cannot last longer than 7.5 minutes under any circumstances, and the average is only 2 to 3 minutes. Totality ends when the Sun's bright surface reappears at the trailing edge of the Moon. Daylight returns quickly, and the corona and chromosphere vanish. This corresponds to the moment when the trailing edge of the Moon's umbra sweeps over the observer.

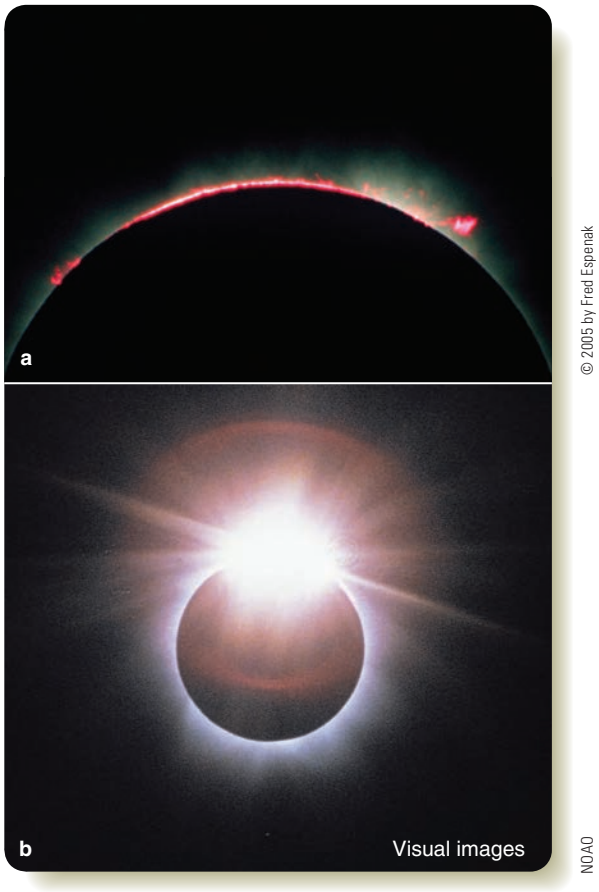
Just as totality begins or ends, a small part of the photosphere can peek through a valley at the edge of the lunar disk. Although it is intensely bright, such a small part of the photosphere does not completely drown out the fainter atmosphere and corona, which forms a silvery ring of light with the brilliant spot of photosphere gleaming like a diamond (Figure 3-12b). This **diamond ring effect** is one of the most spectacular of astronomical sights, but it is not visible during every solar eclipse. Its occurrence depends on the exact orientation and motion of the Moon.

3-3d Observing an Eclipse

Not too many years ago, astronomers traveled great distances to exotic places to get their instruments into the path of totality and study the faint outer corona that is visible only during the few minutes of a total solar eclipse. Now, many of those observations can be made every day by solar telescopes in space, but

A Total Solar Eclipse
The Moon moving from the right just begins to cross in front of the Sun.
The disk of the Moon gradually covers the disk of the Sun.
Sunlight begins to dim as more of the Sun's disk is covered.
During totality, pink prominences are often visible.
A longer-exposure photograph during totality shows the fainter corona.
Visual images
Daniel Good

▲ **Figure 3-11** This sequence of photos shows the first half of a total solar eclipse. The sequence is reversed during the second half of the eclipse. Note that the bottom two frames show the same total eclipse stage with two different photo exposure times.

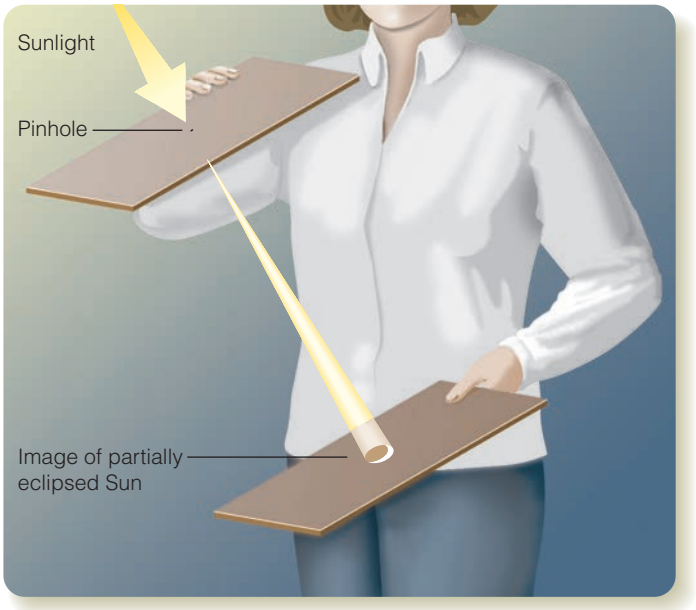


▲ **Figure 3-12** (a) During a total solar eclipse, the Moon covers the photosphere, and the ruby-colored chromosphere and pink prominences are visible. Only the lower corona is visible in this image. (b) The diamond ring effect can sometimes occur momentarily at the beginning or end of totality if a small segment of the photosphere peeks out through a valley at the edge of the lunar disk.

eclipse enthusiasts still journey to remote corners of the world for the thrill of seeing a total solar eclipse.

No matter how thrilling a solar eclipse is, you must be cautious when viewing it. During the partial phase, part of the brilliant photosphere remains visible, so it is hazardous to look at the eclipse without protection. Dense filters do not necessarily provide protection because some filters do not block the invisible infrared (heat) radiation that can burn the retina of your eyes. Dangers like these have led officials to warn the public not to look at solar eclipses at all and have even frightened some people into locking themselves and their children into windowless rooms. It is a **Common Misconception** that sunlight is somehow more dangerous during an eclipse. In fact, it is always dangerous to look at the Sun. The danger posed by an eclipse is that people are tempted to ignore common sense and look at the Sun directly, which can burn their eyes even when the Sun is almost totally eclipsed.

The safest and simplest way to observe the partial phases of a solar eclipse is to use pinhole projection. Poke a small pinhole in a sheet of cardboard. Hold the sheet with the hole in sunlight and allow light to pass through the hole and onto a second sheet of cardboard (**Figure 3-13**). On a day when there is no eclipse, the



▲ **Figure 3-13** A safe way to view the partial phases of a solar eclipse. Use a pinhole in a card to project an image of the Sun on a second card. The greater the distance between the cards, the larger (and fainter) the image will be.

result is a small, round spot of light that is an image of the Sun. During the partial phases of a solar eclipse, the image will show the dark silhouette of the Moon obscuring part of the Sun. Pinhole images of the partially eclipsed Sun can also be seen in the shadows of trees as sunlight peeks through the tiny openings between the leaves and branches. This can produce an eerie effect just before totality as the remaining sliver of Sun produces thin crescents of light on the ground under trees. Once totality begins, it is safe to look directly. The totally eclipsed Sun is fainter than a full moon.

Practicing **SCIENCE**

If you were on Earth watching a total solar eclipse, what would astronauts on the Moon see when they looked at Earth? Answering this question requires that you change your point of view and imagine seeing the eclipse from a new location. Scientists commonly imagine seeing events from multiple points of view as a way to develop and test their understanding.

Astronauts standing on the Moon can see Earth only if they are on the side that faces Earth. Because solar eclipses always happen at new moon, the near side of the Moon would be in darkness, and the far side of the Moon would be in full sunlight. The astronauts on the near side of the Moon would be standing in darkness, and they would be looking at the fully illuminated side of Earth. They would see Earth at full phase. The Moon's shadow would be crossing Earth, and if the astronauts looked closely, they might be able to see the spot of darkness where the Moon's umbral shadow touched Earth. It would take hours for the shadow to cross Earth.

Standing on the Moon and watching the Moon's umbral shadow sweep across Earth would be a cold, tedious assignment. Perhaps you can imagine a more interesting assignment for the astronauts. ***What would astronauts on the Moon see while people on Earth were seeing a total lunar eclipse?***

3-4 Predicting Eclipses

A Chinese legend tells of two astronomers, Hsi and Ho, who were too drunk to predict the solar eclipse of October 22, 2137 BCE. Or perhaps they failed to conduct the proper ceremonies to scare away the dragon that, according to Chinese tradition, was snacking on the Sun's disk. When the emperor recovered from the terror of the eclipse, he had the two astronomers beheaded.

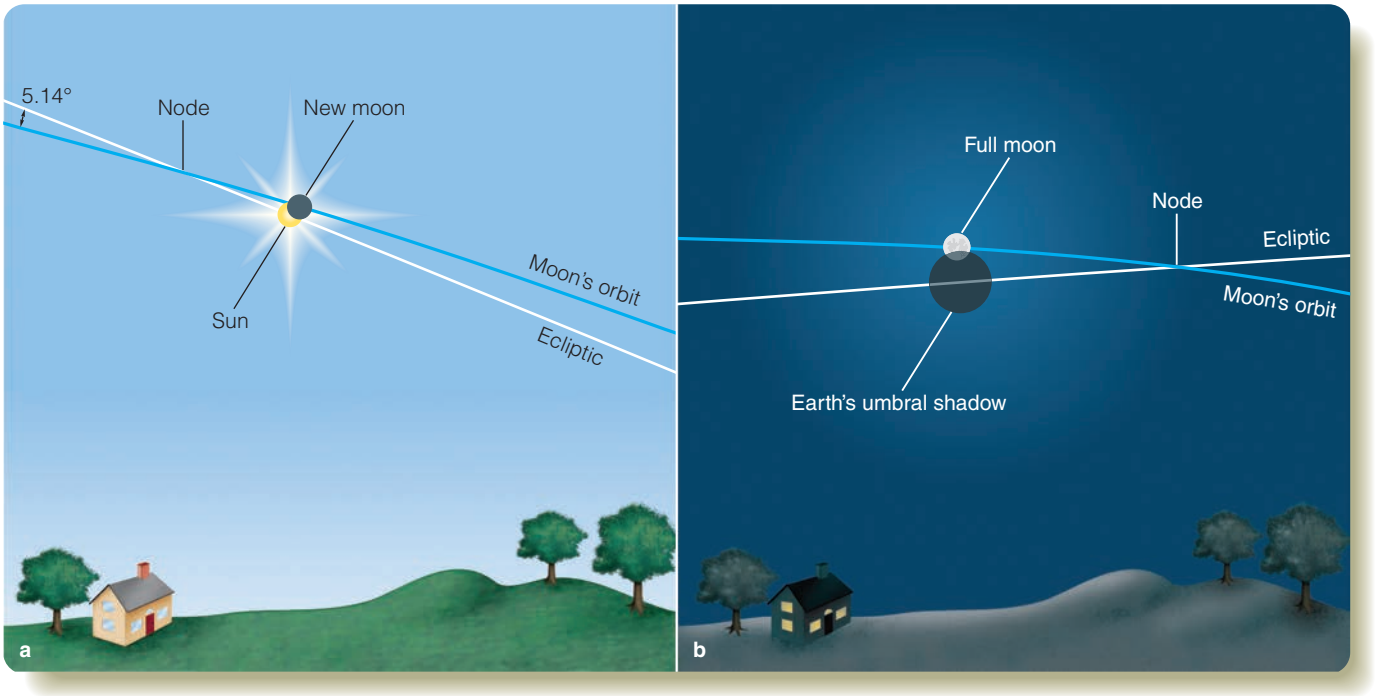
Making exact eclipse predictions requires a computer and proper software, but astronomers in early civilizations could make educated guesses as to which full moons and which new moons might result in eclipses. There are three good reasons to review their methods. First, it is an important chapter in the history of science. Second, it will illustrate how apparently complex phenomena can be analyzed in terms of cycles. Third, eclipse prediction will exercise your scientific imagination and help you visualize Earth, the Moon, and the Sun as objects moving through space.

3-4a Conditions for an Eclipse

You can predict eclipses by thinking about the motion of the Sun and Moon in the sky. Imagine that you can look up into the sky from your home on Earth and see the Sun appearing to move along the ecliptic and the Moon moving along its orbit. Because the orbit of the Moon is tipped slightly more than 5 degrees to the plane of Earth's orbit, you see the Moon follow a path tipped by the same angle to the ecliptic. Each month, the Moon crosses the ecliptic at two points called **nodes**. It crosses at one node going southward, and about two weeks later it crosses at the other node going northward.

Eclipses can occur only when, viewed from Earth, the Sun is near one of the nodes of the Moon's orbit. Only then can the new moon cross in front of the Sun and produce a solar eclipse, as shown in **Figure 3-14a**, and only then can the full moon enter Earth's shadow and be eclipsed. Most new moons pass too far north or too far south of the ecliptic to cause an eclipse (look again at Figure 3-2). (Note that this requirement for eclipses is the reason the Sun's apparent path through the sky is called the *ecliptic*.) Also, when the Sun is near one node, Earth's shadow points near the other node, and a lunar eclipse is possible. A lunar eclipse doesn't happen at every full moon because most full moons pass too far north or too far south of the ecliptic and miss the umbra of Earth's shadow. Some months you might see a partial lunar eclipse, as illustrated in **Figure 3-14b**.

Thus, there are two conditions for an eclipse: The Sun must be near one of the two nodes of the Moon's orbit, and the Moon must pass near either the same node (solar eclipse) or the other node (lunar eclipse). This means, of course, that solar eclipses



▲ **Figure 3-14** Eclipses can occur only when the Sun appears from Earth to be near one of the nodes of the Moon's orbit. (a) A solar eclipse occurs when the Moon meets the Sun near a node. (b) A lunar eclipse occurs when the Sun and Moon are near opposite nodes. Partial eclipses are shown here for clarity.

can occur only when the Moon is new, and lunar eclipses can occur only when the Moon is full.

Now you can understand the ancient secret of predicting eclipses. An eclipse can occur only in a period called an **eclipse season**, during which the Sun is close to a node in the Moon's orbit. For solar eclipses, an eclipse season is about 32 days long. Any new moon during this period will produce a solar eclipse. For lunar eclipses, the eclipse season is a bit shorter, about 22 days. Any full moon in this period will encounter Earth's shadow and be eclipsed.

This makes eclipse prediction easy. All you have to do is keep track of where the Moon crosses the ecliptic (where the nodes of its orbit are). Then, when the Sun approaches either of the nodes you can warn everyone that eclipses are possible. This system works fairly well, and astronomers in early civilizations such as the Maya may have used such a system. You could have been a very successful Mayan astronomer with what you know about eclipse seasons, but you can do even better if you change your point of view.

3-4b The View from Space

Change your point of view and imagine that you are looking at the orbits of Earth and the Moon from a point far away in space. Recall that the Moon's orbit is tipped at an angle to Earth's orbit. The shadows of Earth and Moon are long and thin, as shown in Figure 3-2. That is why it is so easy for them to miss their mark at new moon or full moon and usually fail to produce an eclipse. As Earth orbits the Sun, the Moon's orbit remains approximately fixed in orientation. The nodes of the Moon's orbit are the points where it passes through the plane of Earth's orbit; an eclipse season occurs each time the line connecting these nodes, the **line of nodes**, points directly toward the Sun, allowing the shadows of Earth and Moon then to hit their marks. Look at **Figure 3-15** and notice that the line of nodes does not point at the Sun in the example at lower left, and no eclipses are possible at that time of the year; the shadows miss. At lower right, during an eclipse season, the line of nodes points toward the Sun, and the shadows produce eclipses.

If you watched for years from your point of view in space, you would see the orbit of the Moon precess like a hubcap spinning on the ground. This precession is caused mostly by the gravitational influence of the Sun, and it makes the line of nodes seem to rotate around the sky, as viewed from Earth, once every 18.6 years. As a result, the nodes slip westward along the ecliptic at a rate of 19.4 degrees per year. Consequently, the Sun does not need a full year to go from a node all the way around the ecliptic and appear back at that same node. Because the node is moving westward to meet the Sun, the Sun will cross the node after only 346.6 days (an **eclipse year**). This means that eclipse seasons begin about 19 days

earlier every year (Figure 3-16). If you see an eclipse in late December one year, you can see eclipses in early December the next year, and so on.

Eclipses follow a pattern, and if you were an astronomer in an early civilization who understood the pattern, you could predict eclipses without ever knowing what the Moon is or how an orbit works. Once you have observed a few eclipses from a given location, you know when the eclipse seasons are occurring, and you can predict next year's eclipse seasons by subtracting 19 days. New moons and full moons near those dates are candidates for eclipses.

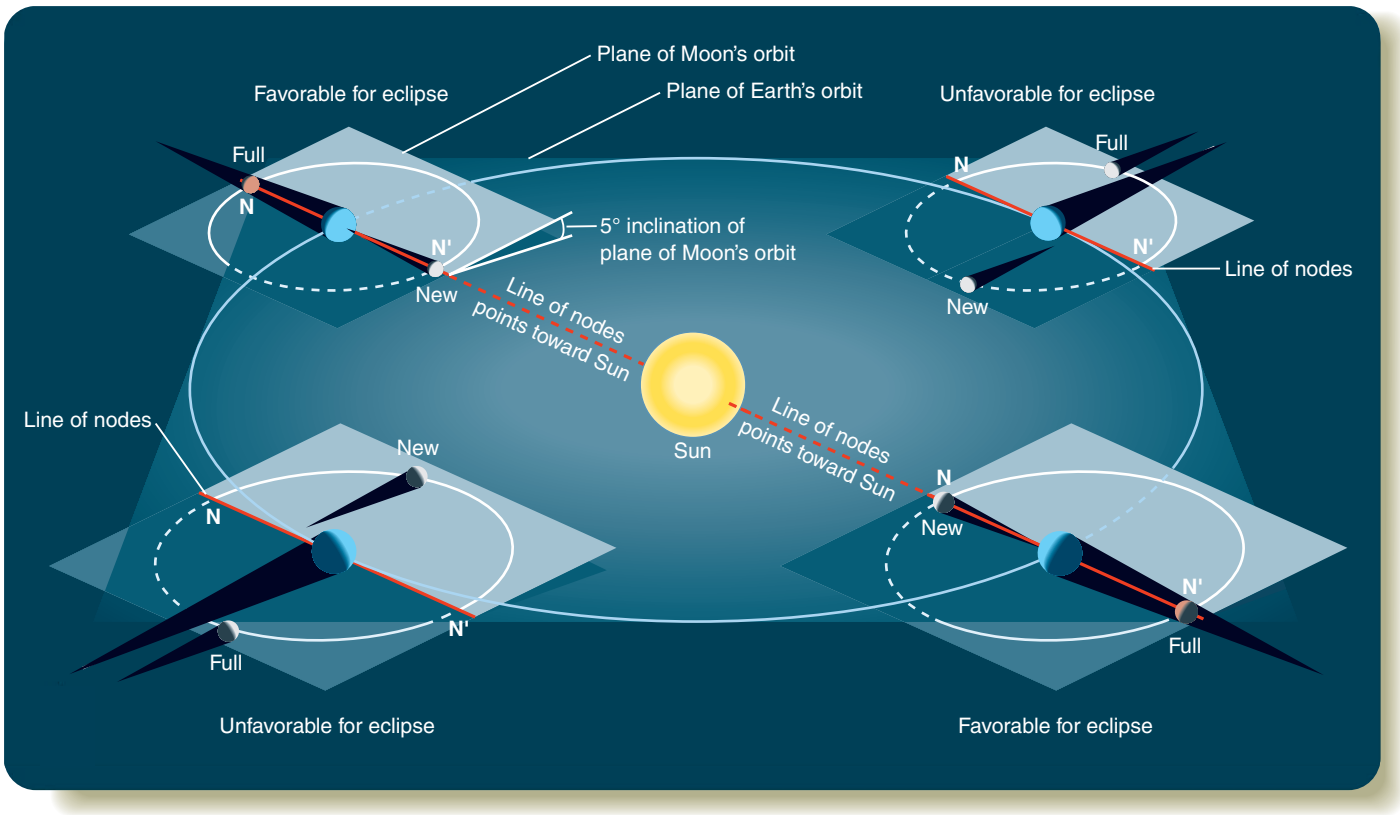
3-4c The Saros Cycle

Astronomers of antiquity could predict eclipses in an approximate way using eclipse seasons, but they could have been much more accurate if they had recognized that eclipses occur following certain patterns. The most important of these is the **Saros cycle** (sometimes referred to simply as the Saros). After one Saros cycle of 18 years 11⅓ days, the pattern of eclipses repeats. In fact, *Saros* comes from a Greek word that means "repetition."

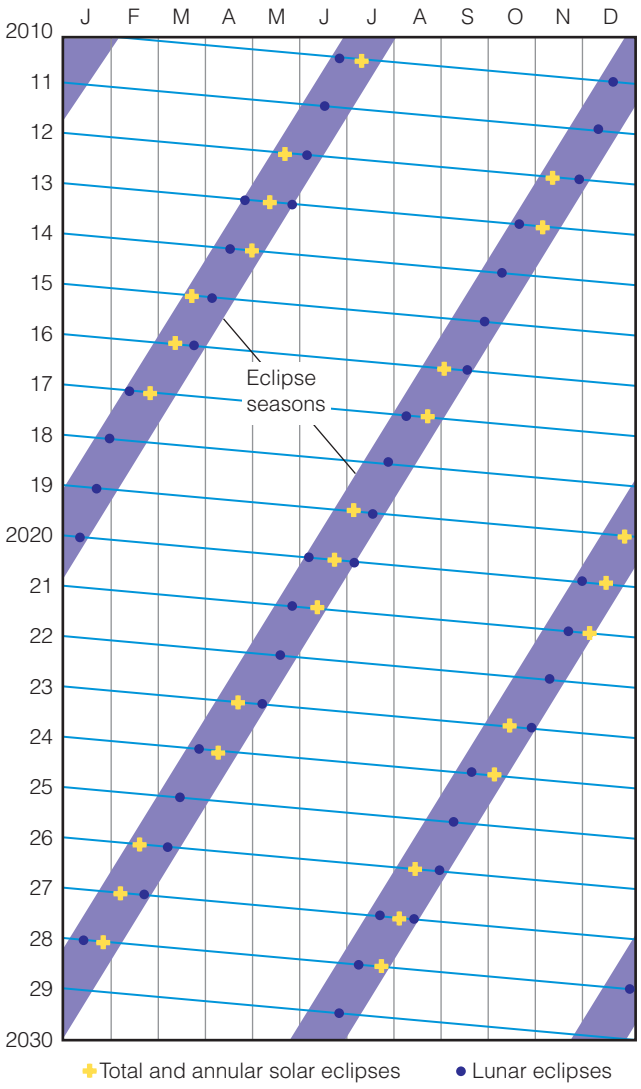
One Saros cycle contains 6585.321 days, which is equal to 223 lunar synodic months. Therefore, after one Saros, the Moon is back to the same phase it had when the cycle began. But, one Saros is also equal to exactly 19 eclipse years. After one Saros cycle, the Sun has returned to the same place it occupied with respect to the nodes of the Moon's orbit when the cycle began. If an eclipse occurs on a given day, then 18 years 11⅓ days later, the Sun, the Moon, and the nodes of the Moon's orbit return to nearly the same relationship, and an eclipse with almost exactly the same geometry (length of totality, general direction of motion of the Moon's shadow, and so on) occurs again.

Although the eclipse geometry repeats almost exactly, it is not visible from the same place on Earth. The Saros cycle is ⅓ of a day longer than 18 years 11 days. When the eclipse happens again, Earth will have rotated ⅓ of a turn farther east, and the eclipse will occur ⅓ of the way westward around Earth (Figure 3-17). That means that after three Saros cycles—a period of 54 years plus 34 days—the same eclipse occurs in about the same part of Earth.

One of the most famous predictors of eclipses was the Greek philosopher Thales of Miletus (about 640–546 BCE), who supposedly learned of the Saros cycle from Babylonian astronomers. No one knows for certain which eclipse Thales predicted, but some scholars suspect it was the eclipse of May 28, 585 BCE. In any case, the eclipse occurred at the height of a battle between the Lydians and the Medes, and the mysterious darkness in midafternoon so alarmed the two armies—thinking they must have offended the gods somehow—that they concluded a truce.

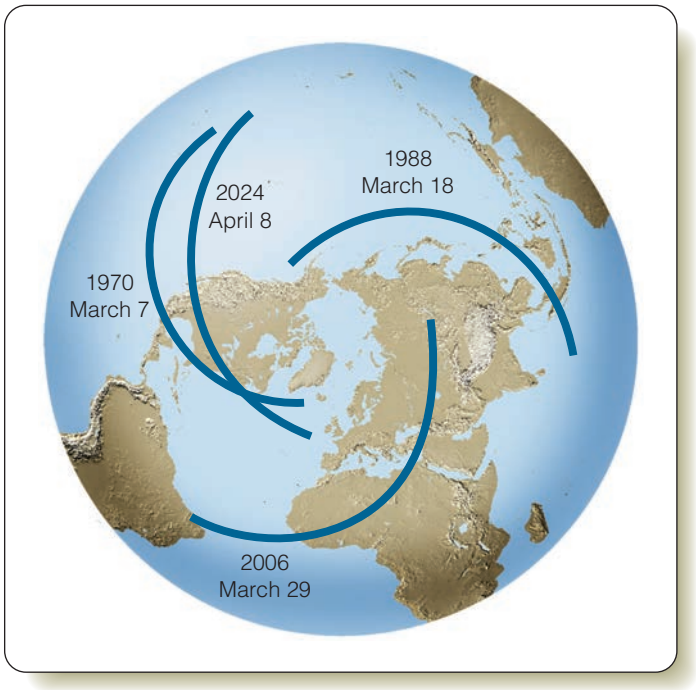


▲ **Figure 3-15** The Moon's orbit is tipped a bit more than 5 degrees to Earth's orbit. The nodes N and N' are the points where the Moon passes through the plane of Earth's orbit. At those parts of Earth's orbit where the line of nodes points toward the Sun, eclipses are possible at new moon and full moon.



▲ **Figure 3-16** A calendar of eclipse seasons: Each year the eclipse seasons begin about 19 days earlier than in the previous year. Any new moon or full moon that occurs during an eclipse season results in an eclipse. Only total and annular eclipses are shown here.

Although there are historical reasons to doubt that Thales actually predicted the eclipse, the important point is that he could have done it. If he had records of past eclipses of the Sun visible from the area, he could have discovered that they tended to recur with a period of 54 years plus 34 days (three Saros cycles). Indeed, he could have predicted the eclipse without having any understanding of the cause of the Saros cycle.



▲ **Figure 3-17** The Saros cycle at work: An eclipse with a track having nearly the same shape as that of the total solar eclipse of March 7, 1970, recurred 18 years $11^1/_3$ days later over the Pacific Ocean. After another interval of 18 years $11^1/_3$ days, an eclipse with a similar track was visible from Asia and Africa. After another 18 years $11^1/_3$ days, in the year 2024, an eclipse like the 1970 eclipse will again be visible from the United States.

Practicing **SCIENCE**

Why can't two successive full moons be totally eclipsed?
Most people suppose that eclipses occur at random or in some pattern so complex you need a big computer to make predictions. In fact, like many natural events, eclipses occur in a cycle that can be observed, and predicting eclipses can be reduced to a series of simple steps.

A lunar eclipse can happen only when the Sun is near one node and the Moon crosses Earth's shadow at the other node. A lunar eclipse season is only 22 days long, and any full moon in that time will be eclipsed. However, the Moon takes 29.5 days to go from one full moon to the next. If one full moon is totally eclipsed, the next full moon 29.5 days later will occur long after the end of the eclipse season, and there can't be a second eclipse.

Now use your knowledge of the cycles of the Sun and Moon to consider a similar question: ***How can the Sun be eclipsed by two successive new moons?***

WHAT ARE WE?

Scorekeepers

The Moon is a companion in our daily lives, in our history, and in our mythology. It makes a dramatic sight as it moves through the sky, cycling through a sequence of phases that has repeated for billions of years. The Moon has been humanity's timekeeper. Moses, Jesus, and Muhammad saw the same Moon that you see. The Moon is part of our human heritage, and famous paintings, poems, plays, and music celebrate the beauty of the Moon.

Eclipses of the Sun and Moon have frightened and fascinated for millennia, and some of humanity's earliest efforts to understand nature were focused on counting the phases of the Moon and predicting eclipses. Some astronomers have found evidence that Stonehenge could have been used for eclipse prediction, and the ancient Maya in Central America left behind elaborate tables that allowed them to predict eclipses.

Our lives are ruled by the Moon as it divides our year into months, and its cycle from new to first quarter to full to third quarter and back to full divides the month into four weeks. In a Native American story, Coyote gambles with the Sun to see if the Sun will return after the winter solstice to warm Earth. The Moon keeps score. The Moon is a symbol of regularity, reliability, and dependability. It is the score-keeper counting out our weeks and months.

Study and Review

SUMMARY

3-1 The Changeable Moon

- The Moon orbits eastward around Earth once a month and rotates on its axis so that the same side faces Earth throughout its orbit.
- Because you see the Moon by reflected sunlight, its appearance changes as it orbits Earth and sunlight illuminates different amounts of the side facing Earth. This repeating pattern of Moon shapes as viewed from Earth is called the **lunar phase** cycle.
- The lunar phases are said to wax from new moon to first quarter to full moon (meaning, the portion of the Moon seen to be illuminated increases) and wane from full moon to third quarter to new moon (illuminated portion decreases).
- A complete cycle of lunar phases takes 29.5 days, which is known as the Moon's **synodic period**. The **sidereal period** of the Moon—its orbital period with respect to the stars—is shorter, about 27.3 days.

3-2 Lunar Eclipses

- If a full moon passes through Earth's shadow, sunlight is cut off, and the Moon darkens in a **lunar eclipse**. If the Moon fully enters the dark **umbra** of Earth's shadow, the eclipse is a **total lunar eclipse;** but if it only grazes the umbra, the eclipse is a **partial lunar eclipse**. If the Moon enters the **penumbra** but not the umbra, the eclipse is a **penumbral lunar eclipse**.
- During **totality**, the eclipsed Moon looks copper red because sunlight refracts through Earth's atmosphere and bounces off the Moon to the night side of Earth.

3-3 Solar Eclipses

- The **small-angle formula** allows you to calculate an object's angular diameter from its linear diameter and distance.

$$\frac{\text{angular diameter (in arc seconds)}}{2.06 \times 10^5} = \frac{\text{linear diameter}}{\text{distance}} \quad \text{(Eq. 3-1)}$$

The angular diameter of the Sun and Moon is about 0.5 degree.

- A **solar eclipse** occurs if a new moon passes between the Sun and Earth and the Moon's shadow sweeps over Earth's surface along the **path of totality**. Observers inside the path of totality see a **total solar eclipse**, and those just outside the path of totality but inside the penumbra see a **partial solar eclipse**.
- When the Moon is near **perigee**—the closest point in its orbit—its angular diameter is large enough to cover the Sun's photosphere and produce a total eclipse. But if the Moon is near **apogee**—the farthest point in its orbit—it looks too small and can't entirely cover the photosphere. A solar eclipse occurring then would be an **annular eclipse**.
- During a total eclipse of the Sun, the bright **photosphere** of the Sun is covered, and the faint low-density **corona**, the **chromosphere**, and **prominences** become visible.
- Sometimes at the beginning or end of the totality phase of a total solar eclipse, a small piece of the Sun's photosphere can peek out through a valley at the edge of the Moon and produce a **diamond ring effect**.
- Looking at the Sun is dangerous and can burn the retinas of your unprotected eyes. The safest way to observe the partial phases of a solar eclipse is by pinhole projection. Only during totality, when the photosphere is completely hidden, is it safe to look at the Sun directly.

3-4 Predicting Eclipses

- Solar eclipses must occur at new moon, and lunar eclipses must occur at full moon. Because the Moon's orbit is tipped a few degrees from the plane of Earth's orbit, most new moons cross north or south of the Sun, and there are no solar eclipses in those months. Similarly, most full moons cross north or south of Earth's shadow, and there are no lunar eclipses in those months.
- The Moon's orbit crosses the ecliptic at two locations called **nodes**, and eclipses can occur only when the Sun and Moon are simultaneously near a node. During these periods, called **eclipse seasons**, a new moon will cause a solar eclipse, and a full moon can have a lunar eclipse. An eclipse season occurs each time the **line of nodes** points toward the Sun. Knowing when the eclipse seasons occur would allow you to guess which new moons and full moons could cause eclipses.

- Because the orbit of the Moon precesses in the retrograde direction, the nodes slip westward along the ecliptic, and it takes the Sun only about 347 days to go from a node around the ecliptic and back to the same node. This length of time is called an **eclipse year**. For that reason, eclipse seasons begin about 19 days earlier each year.
- Eclipses follow a pattern called the **Saros cycle**. After one Saros of 18 years $11^1/_3$ days, the pattern of eclipses repeats. After three Saros, which is 54 years and 34 days, the pattern of eclipses will repeat in approximately the same parts of Earth. Some ancient astronomers knew of the Saros cycle and used it to predict eclipses.

REVIEW QUESTIONS

1. Tonight you see the Moon at midnight at its highest point in your sky. What is the phase of the Moon? Three hours later, what phase will the Moon be in?
2. You are located in Syracuse, NY, United States, the time is 6 PM, and you see the full moon in your clear winter night sky. You call your aunt in San Diego, CA, United States, and ask her to go outside to view the Moon with you. In which direction are you looking to see the full moon: north, south, east, or west? Your aunt tells you that the weather in San Diego is clear as she is stepping outside. What does your aunt report about the Moon's phase and location in her sky?
3. You are located in Knoxville, TN, United States. Your friend is located in Lima, Peru. You see a waning gibbous in your clear night sky. What phase, if any, will your friend see if the night sky in Lima is also clear?
4. Which lunar phases would be visible in the sky at dawn? At midnight?
5. Tonight you see a waxing crescent moon. Seven days from now, which phase will you see if the night sky is clear?
6. You look along the easterly horizon and see a crescent moon rising in the clear night sky. What time is it? Which side—right or left—of the Moon's near side is illuminated?
7. If you looked back at Earth from the Moon, what phase would Earth have when the Moon was full? New? At first quarter? At waxing crescent?
8. The phase of the Moon is a waning crescent as viewed from Earth. You are located in the dark portion of the Moon's near side, near the Moon's equatorial region. Which side of Earth from your vantage point on the Moon—the left, the right, the top, or the bottom side—is illuminated by the Sun?
9. If a planet has a moon, must that moon go through the same phases that Earth's Moon displays?
10. Could a solar powered spacecraft generate any electricity while passing through Earth's umbral shadow? Through Earth's penumbral shadow?
11. If a lunar eclipse occurred at midnight, where in the sky would you look to see it?
12. If Earth had no atmosphere, what color would the eclipsed moon appear in the sky?
13. If the Moon orbited Earth from North Pole to South Pole instead of near the ecliptic, would lunar and solar eclipses still occur? Would the moon phase still have to be full or new?
14. Why do solar eclipses happen only at new moon? Why not every new moon?
15. Why isn't the corona visible during partial or annular solar eclipses?
16. Which has the larger angular diameter in the sky—the Sun or Moon—during an annular eclipse? If you wanted to be in the umbra, where would you have to physically be located to see this annular eclipse as a total solar eclipse?
17. What is the angular diameter of the Moon in the third-quarter phase? What is the shortest/longest angular distance from the horizon to the Moon in the third-quarter phase when the time is midnight or noon?
18. Why can't the Moon be eclipsed when it is halfway between the nodes of its orbit?
19. Why are solar eclipses separated by one Saros cycle not visible from the same location on Earth?
20. How could Thales of Miletus have predicted the date of a solar eclipse without observing the location of the Moon in the sky?
21. Will an eclipse occur in February 2025? In July 2028? If so, what kind?
22. **How Do We Know?** Some people think science is like a grinder that cranks data into hypotheses. What would you tell them about the need for scientists to be creative and imaginative?

ACTIVE INQUIRY QUESTIONS

1. How would eclipses be different if the Moon's orbit were not tipped with respect to the ecliptic, the plane of Earth's orbit?
2. You happen to visit the Moon when some people on Earth see a total solar eclipse. Who has a more spectacular sight during this event, you or the friends you left behind back on Earth? Why?

PROBLEMS

1. Pretend the Moon's orbit around Earth is a perfect circle. How long does it take in units of days for the Moon to move 90 degrees relative to the stars? Is this number dependent on the synodic period or the sidereal period?
2. Identify the phases of the Moon if on March 20 the Moon is located at the point on the ecliptic called: (a) the vernal equinox, (b) the autumnal equinox, (c) the summer solstice, (d) the winter solstice.
3. Identify the phases of the Moon if at sunset in the Northern Hemisphere the Moon is: (a) near the eastern horizon, (b) high in the southern sky, (c) in the southeastern sky, (d) in the southwestern sky.
4. What fraction of the Moon's surface area is the far side? Of the near side of a third-quarter moon, what fraction is dark? What fraction of the far side is in the dark that cannot be seen by an observer from Earth viewing the Moon in its third-quarter phase?
5. About how many days must elapse between first-quarter moon and third-quarter moon in the same cycle?
6. Tonight you see a waning crescent in the night sky. A few days later, the night is once again clear and you see a waning crescent. How many degrees did the Moon advance in its orbit during this time frame?
7. If on March 1 the Moon is full and is near Famous Star Spica, when will the Moon next be near Spica? When will it next be full? Are these values the same? Why or why not?
8. How many times larger than the Moon is the diameter of Earth's umbral shadow at the Moon's distance? (*Hint:* See the photo in Figure 3-4.)
9. Use the small-angle formula to calculate the angular diameter of Earth as seen from the Moon. (*Note:* The linear diameter of Earth is 12,700 km.)
10. Use the small-angle formula (Eq. 3-1) to calculate the angular diameter of the Sun as seen from Earth if the Sun were at the location of the Moon. Show your answer in units of degrees. (*Note:* The linear diameter of the Sun is given in this chapter.)
11. At perigee, the Moon is closer than average by 21,100 km. At apogee, the Moon is further than average by 21,100 km. Is the angular diameter more or less at perigee than apogee? What is the angular diameter of the Moon at perigee? At apogee? By how much greater a percentage is the angular diameter larger or smaller at perigee than at the average distance? At apogee? (*Hint*: The Moon's average distance from Earth is given in this chapter.)
12. Examine the list of upcoming lunar eclipses in Table 3-1. What fraction of years have two eclipses?

13. During solar eclipses, large prominences are often seen extending as much as 5 arc minutes from the edge (limb) of the Sun's disk. How far is that in kilometers? In Earth diameters? If you used protective glasses, do you think you could see prominences around the Sun's limb? Why or why not? (*Hints*: Use the small-angle formula, Eq. 3-1. The Sun's average distance is given in this chapter. Earth's diameter is given in Problem 9.)
14. If a solar eclipse occurs on October 3: (a) Why can't there be a lunar eclipse on October 13 of that same year? (b) Why can't there be a solar eclipse on December 28 of that same year?
15. A total eclipse of the Sun was visible from Canada on July 10, 1972. When did an eclipse occur next with the same Earth–Moon–Sun geometry? From what part of Earth was it total?
16. When will the eclipse described in Problem 15 next be total as seen from Canada?
17. When will the eclipse seasons occur during the current year? How many total of all types will occur? Which type of eclipse(s) will occur?
18. Examine Figure 3-16. List the letter *S* for each total or annular solar eclipse that occurs from July 2019 through July 2028 in chronological order. When only a lunar eclipse occurs, put *N* for the word *none*. Do you see a pattern? If so, identify it and predict the next two letters.

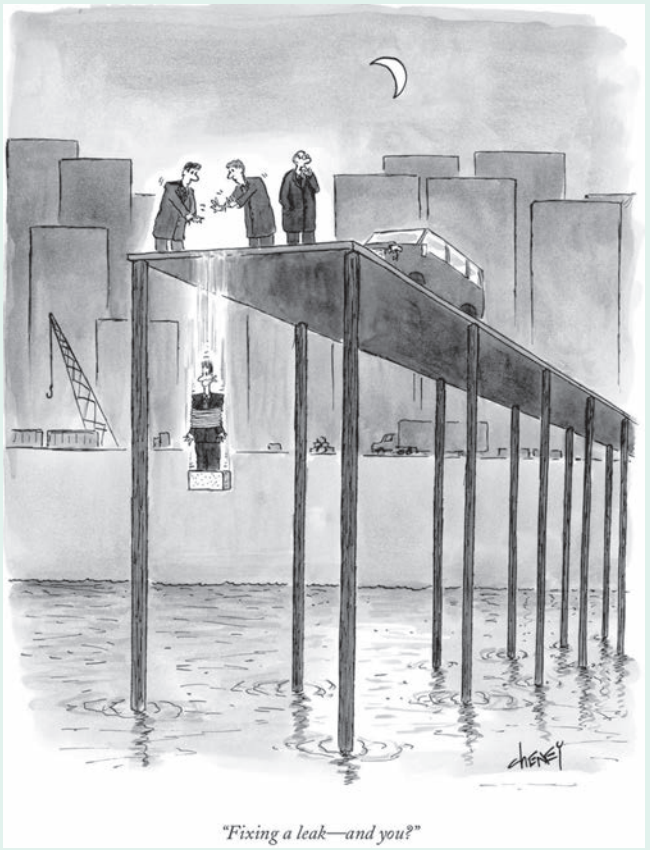
SENSE OF PROPORTION

1. Arrange the following in order of increasing length: one year; one Saros cycle; one revolution of the Moon around Earth; a lunar eclipse from start to finish; totality as seen from a single location on Earth during a solar eclipse
2. Which of the following are roughly the same angular diameter as seen in the sky? the Sun; the bowl of the Big Dipper; Polaris; a full moon

LEARNING TO LOOK

1. Look at the Chapter 2 Concept Art: **The Sky Around You**. What phase of the Moon is the woman viewing?
2. To take the photos that are combined on the opening page of this chapter, was the photographer located on the day, or night, side of Earth? Was the photographer in Earth's umbra, penumbra, or both?
3. Look at the Chapter 3 Concept Art: **The Phases of the Moon**. Find the person looking at the third-quarter phase of the Moon at sunrise. What percentage of the near side of the Moon is illuminated? Likewise, what percentage is in the dark? Repeat the exercise for the new phase of the Moon.
4. Look at the Chapter 3 Concept Art: **The Phases of the Moon**. Find the waxing gibbous phase of the Moon. If this phase could be seen at its highest point in your winter sky, is it daytime or nighttime? Approximately what time is it? At approximately what time did that phase rise over the eastern horizon? At approximately what time will that phase set over your western horizon? Repeat the exercise for the waning crescent phase.
5. Use the photos in Figure 3-1 as evidence to show that the Moon always keeps the same side facing Earth.
6. Draw the umbral and penumbral shadows onto the diagram in Figure 2 of **The Phases of the Moon** concept art. Use the diagram to explain why lunar eclipses can occur only at full moon and solar eclipses can occur only at new moon.
7. Look at Figure 3-4. What phase of the Moon is being viewed? Are you looking at the near side or far side? Are you on the daytime or nighttime side of Earth? Are you in the umbra or penumbra, and which celestial object is casting the shadow?
8. Do you think the color of the chromosphere in Figure 3-12a is falsely colored ruby red? Why or why not?
9. What evidence of the Saros cycle can you see in Figure 3-16?

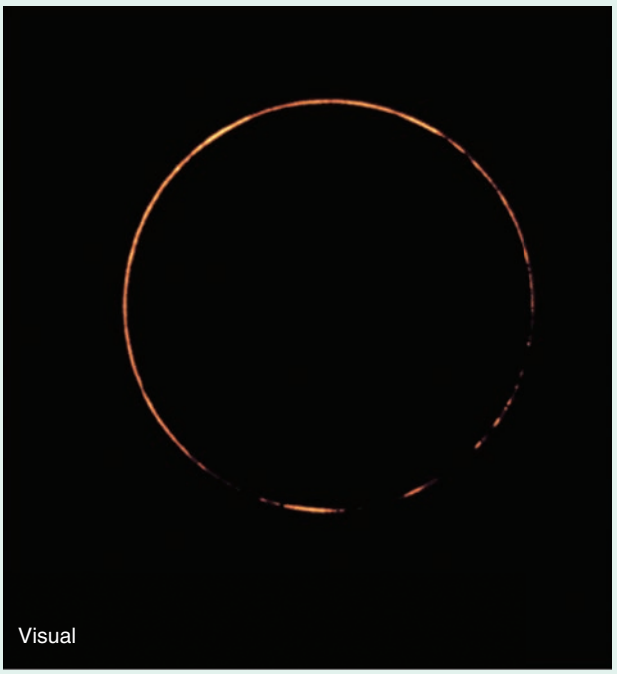
10. The accompanying cartoon shows a crescent moon. Explain why the Moon could never look this way at night.



"Fixing a leak—and you?"

© The New Yorker Collection 2006 Tom Cheney from cartoonbank.com. All Rights Reserved

11. This photo shows the annular eclipse of May 30, 1984. How is it different from the annular eclipse shown in Figure 3-10c?



Laurence Marschall

4 Origins of Modern Astronomy

Galileo explaining moon topography to skeptics (colour litho)/Huens, Jean-Leon (1921-82)/NATIONAL GEOGRAPHIC SOCIETY/Bridgeman Images

▲ Galileo's telescope revealed such things as craters on the Moon, phases of Venus, and the existence of four moons orbiting Jupiter. He demonstrated his telescope and discussed his observations with powerful people, explaining how observational evidence could be used to test the prevailing Earth-centered model of the Universe.

The preceding three chapters gave you a modern view of the ways in which Earth, the Moon, and the Sun move through space, and how those motions produce the sights you see in the sky. But how did humanity first realize that we live on a planet moving through space? That required the overthrow of ancient and honored ideas of Earth's place in the Universe.

In this chapter, you will discover how Renaissance scientists such as Copernicus, Kepler, and Galileo proposed new models and changed the way we understand nature. Here you will find answers to four important questions about the transition from ancient to modern views of the Universe:

- How did classical philosophers describe Earth's place in the Universe?
- How did Copernicus revise those ancient ideas?
- How did Kepler discover the laws of planetary motion?
- How did Galileo's observations support the Copernican model?

This chapter is not just about the history of astronomy. As the astronomers of the Renaissance struggled to understand Earth, the Solar System, and the Universe, they invented a new way of understanding nature—a way of thinking that is now called *science*. Every chapter that follows will use the methods that were invented during a single century that spanned the careers of Copernicus, Kepler, and Galileo.

> **How you would burst out laughing, my dear Kepler, if you would hear what the greatest philosopher of the Gymnasium told the Grand Duke about me…**
>
> FROM A LETTER BY GALILEO GALILEI

FOUR CENTURIES AGO, Galileo was condemned by the Inquisition for his part in a huge controversy about the nature of the Universe, a controversy that focused on two problems. The place of Earth was the most acrimonious issue: Was it the center of the Universe, or was the Sun at the center? A related issue was the nature of planetary motion. Ancient astronomers could see the Sun, Moon, and planets moving along the ecliptic, but they could not describe or predict those motions precisely. To understand the place of Earth in the Universe, astronomers first had to understand planetary motion.

4-1 Roots of Astronomy

Astronomy has its origin in a noble human trait: curiosity. Just as modern children ask their parents what the stars are and why the Moon has phases, early humans asked themselves those same questions. Their answers, which were often couched in mythical or religious terms, reveal great reverence for the order of the heavens.

4-1a Archaeoastronomy

Most of the history of astronomy is lost forever. You can't go to a library or search the Internet to find out what the first astronomers thought about the world because they left no written records. The study of the astronomy of ancient peoples, called **archaeoastronomy** (a combination of "archaeology" and "astronomy"), yields abundant evidence that seeking to understand the heavens is part of human nature.

Perhaps the best-known object investigated by archaeoastronomy is also a major tourist attraction. Stonehenge, standing on Salisbury Plain in southern England, was built in stages from about 3100 BCE to about 1600 BCE, a period extending from the late Stone Age into the Bronze Age. Though the public is most familiar with the monument's massive stones, they were added late in its history. During its first stages, Stonehenge consisted of a circular ditch slightly larger in diameter than the length of a U.S. football field, with a concentric bank just inside the ditch and a long avenue leading away toward the northeast. A massive stone, the Heelstone, stood then, as it does now, outside the ditch in the opening of the avenue.

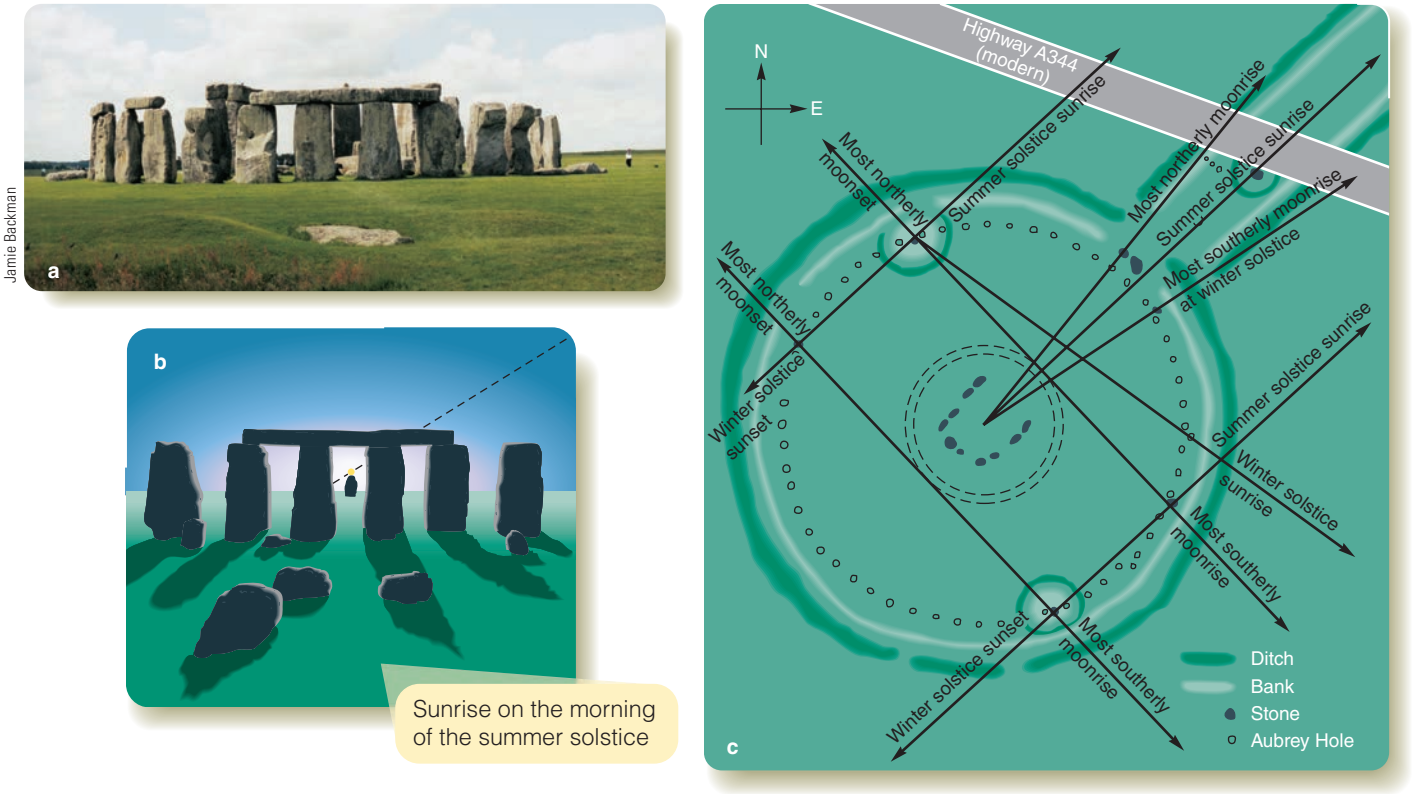
As early as 1740, the English scholar William Stukely suggested that the avenue pointed toward the rising Sun at the summer solstice, but few historians accepted that it was intentional. Nevertheless, seen from the center of the monument, the summer solstice sun does rise behind the Heelstone. More recently, astronomers have recognized other significant astronomical alignments at Stonehenge. For example, there are sight lines that point toward the most northerly and most southerly horizon rising points of the Moon (**Figure 4-1**).

The significance of these alignments has been debated. Some have claimed that the Stone Age people who built Stonehenge were using it as a device to predict lunar eclipses. After studying eclipse prediction in the previous chapter, you understand that predicting eclipses is easier than most people realize, so perhaps it was used in that way, but the truth may never be known. The builders of Stonehenge had no written language and left no records of their intentions. Nevertheless, the presence of solar and lunar alignments at Stonehenge and at many other Stone Age monuments dotting England and continental Europe shows that so-called primitive peoples were paying close attention to the sky. Building astronomical alignments into structures gives the structures special, even holy, meaning by connecting them with the heavens. The roots of astronomy lie not in sophisticated science and mathematics but in human curiosity and awe.

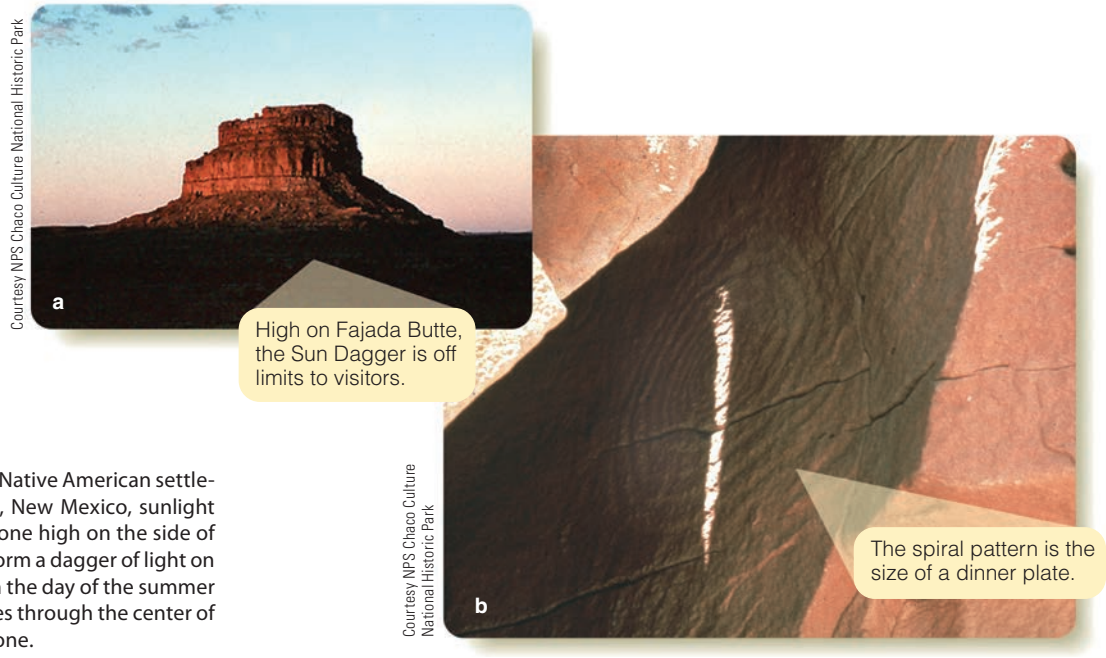
Astronomical alignments in sacred structures are common all around the world. For example, many tombs are oriented toward the rising Sun, and Newgrange, a 5000-year-old passage grave in Ireland, faces southeast so that, at dawn on the day of the winter solstice, light from the rising Sun shines into its long passageway and illuminates the central chamber. No one today knows what the alignment meant to the builders of Newgrange. Whatever its original purpose, Newgrange is clearly a sacred site linked by its alignment to the order and power of the sky.

Some alignments may have served purposes related to keeping an annual calendar. The 2000-year-old Temple of Isis in Dendera, Egypt, was built to align with the rising point of the bright star Sirius. Each year, the first appearance of this star in the dawn twilight marked the flooding of the Nile, so it was an important date indicator. The link between Sirius and the Nile was described in Egyptian mythology; the goddess Isis was associated with the star Sirius, and her husband, Osiris, was linked to the constellation now called Orion, and also to the Nile, the source of Egypt's agricultural fertility.

An intriguing American site in New Mexico, known as the Sun Dagger, unfortunately has no surviving mythology to tell its story. At noon on the day of the summer solstice, a narrow dagger of sunlight shines across the center of a spiral carved on a cliff face high above the desert floor (**Figure 4-2**). The purpose of the Sun Dagger is open to debate, but similar examples have been found throughout the U.S. Southwest. It may have had more of a symbolic and ceremonial purpose than a precise calendar function. In any case, it is just one of the many



▲ **Figure 4-1** (a) The central horseshoe of upright stones is only the most obvious part of Stonehenge. (b) The best-known astronomical alignment at Stonehenge is the summer solstice sun rising over the Heelstone. (c) Although a number of astronomical alignments, indicated on the diagram, have been found at Stonehenge, experts debate their significance.


Courtesy NPS Chaco Culture National Historic Park
a
High on Fajada Butte, the Sun Dagger is off limits to visitors.
Courtesy NPS Chaco Culture National Historic Park
b
The spiral pattern is the size of a dinner plate.

▲ **Figure 4-2** (a) In the ancient Native American settlement known as Chaco Canyon, New Mexico, sunlight shines between two slabs of stone high on the side of 440-foot-high Fajada Butte to form a dagger of light on the cliff face. (b) About noon on the day of the summer solstice, the dagger of light slices through the center of a spiral pecked into the sandstone.

astronomical alignments that ancient people built into their structures to link themselves with the sky.

Some archaeoastronomers study small artifacts made thousands of years ago rather than large structures. Scratches on certain bone and stone implements follow a pattern that may record the phases of the Moon (Figure 4-3). Although controversial, such finds suggest that some of the first human written records were of astronomical phenomena.

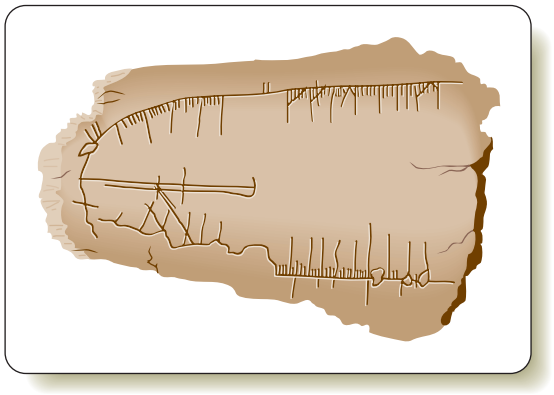
Archaeoastronomy research uncovers the earliest roots of astronomy and, in so doing, reveals some of the first human efforts at systematic inquiry. The most important lesson of archaeoastronomy is that humans don't have to be technologically advanced to have a sophisticated understanding of celestial cycles.

Although the methods of archaeoastronomy can show how ancient people observed the sky, their thoughts about the Universe are, in many cases, unknown. Many cultures had no written language. In other cases, the written record has been lost or even intentionally destroyed. For instance, dozens, possibly hundreds, of beautiful Mayan manuscripts were burned by Spanish missionaries who thought these were the work of the Devil. Only four of the Mayan books survived, and all four include astronomical references. One contains complicated tables that allowed the Maya to predict the motion of Venus and eclipses of the Moon. No one will ever know the extent of what was lost.

The fate of the Mayan books illustrates one reason why histories of astronomy usually begin with the Greeks. Our culture and language is partly descended from theirs, and some of their writing has survived. From that, you can discover what the Greeks thought about the geometry and motions of the heavens.

4-1b The Astronomy of Classical Greece

Greek astronomy was derived from Babylon and Egypt, but the Greek philosophers took a new approach. Rather than relying on religion and astrology, the Greeks proposed a rational universe whose secrets could be understood through logic and reason.



▲ **Figure 4-3** A fragment of a mammoth tusk found in Ukraine with an age of at least 15,000 years contains scribe marks on its edge, simplified in this drawing. These markings have been interpreted as a record of four cycles of lunar phases.

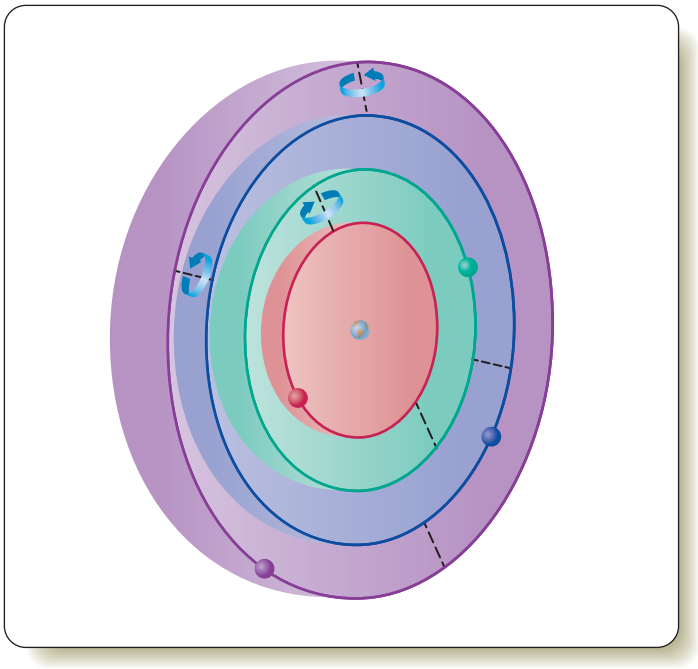
As you study early Greek astronomy, you should keep in mind that the Greeks knew much less about the Universe than you do. As you learned or were reminded of in Chapter 1, the stars are other suns scattered through space within the galaxy called the Milky Way. You also realize that out to the greatest distances the largest telescopes can see, the Universe is filled with other galaxies. Early astronomers knew none of this. They saw the Sun, Moon, and planets moving in regular patterns against a background of stars, and they imagined that the entire Universe was hardly more than the bright objects in the Solar System, extending just a little way beyond the planets to an enclosing sphere carrying the stars. In most cases, they believed we inhabit a **geocentric universe**, with Earth at the center. You will see in this chapter that a small number of astronomers proposed a **heliocentric universe** with the Sun at the center.

When classical Greek astronomers analyzed the heavens using logic and reason, they took the first step toward modern science, which was made possible especially by two early Greek philosophers. Thales of Miletus (c. 624–c. 546 BCE) lived and worked in what is now Turkey. He taught that the Universe is rational and that the human mind can understand why the Universe works the way it does. This view contrasts sharply with that of earlier cultures, which believed that the ultimate causes of things are mysteries beyond human understanding. To Thales and his followers, the mysteries of the Universe were mysteries only because they were unknown, not because they were unknowable.

The second philosopher who made the new scientific attitude possible was Pythagoras (c. 570–c. 495 BCE). He and his students noticed that many things in nature seem to be governed by geometrical or mathematical relations. Musical pitch, for example, is related in a regular way to the lengths of plucked strings. This led Pythagoras to propose that all nature was underlain by musical principles, by which he meant mathematics. One result of this philosophy was the later belief that the harmony of celestial movements produced actual music that was called "the music of the spheres." At a deeper level, the teachings of Pythagoras made Greek astronomers look at the Universe in a new way. Thales said that the Universe could be understood, and Pythagoras said that the underlying rules were mathematical.

In trying to understand the Universe, Greek astronomers did something that Babylonian astronomers had never done: They tried to describe the Universe using geometrical forms. Philolaus (5th century BCE) hypothesized that Earth moved in a circular path around a central fire (not the Sun), which was always hidden behind a counter-Earth located between the fire and Earth. This was the earliest known example of a hypothesis that Earth is in motion.

The great philosopher Plato (c. 424–347 BCE) was not an astronomer, but his teachings influenced astronomy for 2000 years. Plato argued that the reality humans see is only a distorted shadow of a perfect, ideal form. If human observations are distorted, then



▲ **Figure 4-4** The spheres of Eudoxus explain the motions in the heavens by means of nested spheres rotating about various axes at different rates. Earth is located at the center. In this illustration, only 4 of the 27 spheres are shown.

observation can be misleading, and the best path to truth, said Plato, is through pure thought on the ideal forms that underlie nature.

Plato agreed with other philosophers on a principle known as the *perfection of the heavens.* They saw the beauty of the night sky and the regular motion of the heavenly bodies and therefore concluded that the heavens represented perfection. Plato argued that the most perfect geometrical form was the sphere; therefore, the perfect heavens must be made up of spheres rotating at constant rates and carrying objects around in circles. Consequently, later astronomers tried to describe the motions of the heavens by imagining multiple rotating spheres. This became known as the principle of **uniform circular motion**.

Eudoxus of Cnidus (408–355 BCE), a student of Plato, applied this principle when he devised a system of 27 nested spheres that rotated at different rates about different axes to produce a mathematical description of the motions of the Universe (**Figure 4-4**).

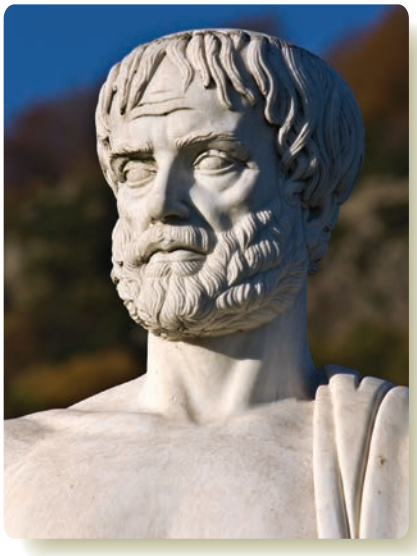
At the time of the Greek philosophers, it was common to refer to systems such as that of Eudoxus as descriptions of the world, where the world included not only Earth but all of the heavenly spheres; today, you would say they were attempting to describe the Universe. The reality of these spheres was open to debate. Some thought of the spheres as nothing more than mathematical ideas that described motion in the world model, whereas others began to think of the spheres as real objects made of perfect celestial material. Aristotle, for example, seems to have thought of the spheres as real.

4-1c Aristotle and the Nature of Earth

Aristotle (384–322 BCE), another of Plato's students, made his own unique contributions to philosophy, history, politics, ethics, poetry, drama, and other subjects (**Figure 4-5**). Because of his far-ranging and penetrating insights, he became the greatest authority of antiquity, and later philosophers referred to him as "The Philosopher." His astronomical model was accepted with only minor alterations for almost 2000 years.

Much of what Aristotle wrote about scientific subjects was wrong, but that is not surprising. The modern scientific method, with its insistence on evidence and hypothesis, had not yet been invented. Aristotle, like other philosophers of his time, attempted to understand his world by reasoning logically and carefully from **first principles**. A first principle is something that is held to be obviously true and needs no further examination. The perfection of the heavens, for instance, was, for Aristotle, a first principle. Once a principle is recognized as true, whatever can be logically derived from it must also be true.

Aristotle believed that the Universe was divided into two parts: Earth, imperfect and changeable, and the heavens, perfect and unchanging. Like most of his predecessors, he believed that Earth was the center of the Universe, so his model is a geocentric universe. The heavens surrounded Earth, and he added parts to the model proposed by Eudoxus to bring the total to 55 crystalline spheres turning at different rates and at different angles to carry the Sun, Moon, and planets across the sky with their observed motions. The lowest sphere, that of the Moon, marked the boundary between the changeable imperfect region containing Earth and the unchanging perfection of the celestial realm beyond the Moon.



PANAGIOTIS KARAPANAGIOTIS/Alamy Stock Photo

▲ **Figure 4-5** Aristotle wrote on such a wide variety of subjects and with such deep insight that he became the great authority on all matters of learning. His opinions on the nature of Earth and the sky were widely accepted for almost two millennia.

Because he believed Earth to be unmoving, Aristotle had to assume that the entire nest of spheres whirls westward around Earth every 24 hours to produce day and night. Different spheres had to move at different rates to produce the motions of the Sun, Moon, and planets against the background of the stars. Because his model was geocentric, he taught that Earth could be the only center of motion, meaning that all of the whirling spheres had to be centered on Earth.

About a century after Aristotle, the Alexandrian philosopher Aristarchus proposed that Earth rotates on its axis and revolves around the Sun. These ideas are, of course, generally correct, but most of the writings of Aristarchus were lost, and his theory was not well known to subsequent scholars. In fact, virtually all later astronomers rejected any suggestion that Earth could move not only because they could feel no motion but also because it conflicted with the teachings of the great Aristotle.

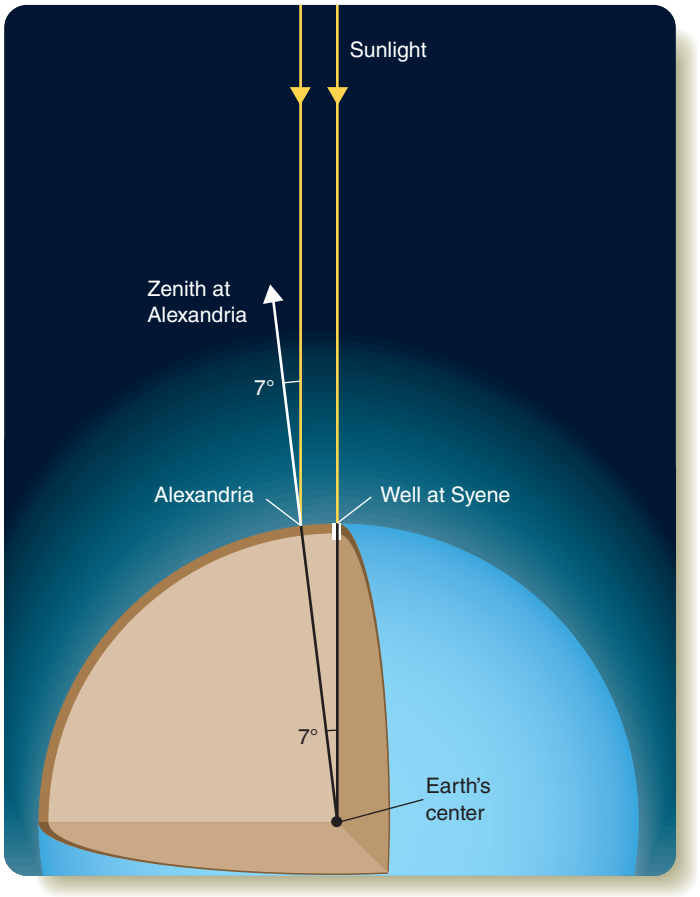
Aristotle taught that Earth must be a sphere because it always casts a round shadow during lunar eclipses, but he could only roughly estimate its size. About 200 BCE, Eratosthenes (c. 276–c. 195 BCE), working in the great library in the Egyptian city of Alexandria, found a way to measure Earth's radius. He learned from travelers that the city of Syene (Aswan) in southern Egypt contained a well into which sunlight shone vertically on the day of the summer solstice. This told him that the Sun was at the zenith at Syene, but on that same day in Alexandria, he noted that the Sun was 1/50 of the circumference of the sky (about 7 degrees) south of the zenith.

Because sunlight comes from such a great distance, its rays arrive at Earth traveling almost parallel. That allowed Eratosthenes to use simple geometry to conclude that the distance from Alexandria to Syene is 1/50 of Earth's circumference (**Figure 4-6**).

To find Earth's circumference, Eratosthenes had to learn the distance from Alexandria to Syene. Travelers told him it took 50 days to cover the distance, and he knew that a camel can travel about 100 stadia per day. This meant the total distance was about 5000 stadia. If 5000 stadia is 1/50 of Earth's circumference, then Earth must be 250,000 stadia in circumference, and, dividing by 2π, Eratosthenes found Earth's radius to be about 40,000 stadia.

How accurate was Eratosthenes's estimate? The stadium (singular of stadia) had different lengths in ancient times. If Eratosthenes used the Olympic stadium, his result was 14 percent bigger than the true value—not bad at all. This was a much better measurement of Earth's radius than Aristotle's estimate, which was much smaller, about 40 percent of the true radius.

You might think this is just a disagreement between two ancient philosophers, but it is related to a **Common Misconception**. Christopher Columbus did not have to convince Queen Isabella that the world was round. At the time of Columbus, all educated people (a small fraction of the population) knew that the world was round and not flat, but they weren't sure how big it was. Columbus, like many others, adopted Aristotle's diameter for Earth, so he thought Earth was small enough that he could sail west and reach Japan and the Spice Islands of the East Indies in a couple of months. If he had accepted Eratosthenes's diameter, Columbus would never have risked the voyage. He and his crew were lucky that America was in the way because if there had been open ocean all the way to Japan, they would have starved to death long before they reached land.



▲ **Figure 4-6** On the day of the summer solstice, sunlight fell to the bottom of a well at Syene, but on the same day the Sun was about 1/50 of a circle (7 degrees) south of the zenith at Alexandria. From this, Eratosthenes was able to calculate Earth's radius.

Aristotle, Aristarchus, and Eratosthenes were philosophers, but the next person you will meet was a real astronomer who observed the sky in detail. Little is known about Hipparchus, who lived about two centuries after Aristotle, during the 2nd century BCE. He is usually credited with the invention of trigonometry, the creation of the first star catalog, and the discovery of precession (look back to Figure 2-7). Hipparchus also described the motion of the Sun, Moon, and planets as following circular paths with Earth near—but not at—their centers. These off-center circles are now known as **eccentrics**. Hipparchus recognized that he could reproduce this motion in a model where each celestial body traveled around a small circle that followed a larger

circle around Earth. The compounded circular motion that he devised became the key element in the masterpiece of the last great astronomer of classical times, Claudius Ptolemaeus, who is better known as Ptolemy (pronounced *TAHL-eh-mee*; the initial *P* is silent) and whom you met in Chapter 2.

4-1d The Ptolemaic Universe

Ptolemy was one of the great astronomer–mathematicians of antiquity. His nationality and birth year are unknown, but around the year 140 he lived and worked in the Greek settlement at Alexandria in what is now Egypt. He ensured the continued acceptance of Aristotle's universe by transforming it into a sophisticated mathematical model.

When you study **An Ancient Model of the Universe**, notice three important ideas and five new terms that show how first principles influenced early descriptions of the Universe and its motions:

1. Ancient philosophers and astronomers accepted as first principles that the heavens were geocentric with Earth located at the center and the Sun, Moon, and planets moving in uniform circular motion. It seemed clear to them that Earth was not moving because they saw no *parallax* in the positions of the stars.
2. The observed motion of the planets did not fit the theory very well. The apparent *retrograde motion* of the planets was difficult to explain with a model of an unmoving Earth at the center of the Universe plus uniform circular motion of celestial objects.
3. In his book that later came to be known as the *Almagest,* Ptolemy attempted to explain the motion of the planets by devising a small circle, an *epicycle,* that rotated along the edge of a larger circle, the *deferent,* which enclosed a slightly off-center Earth. An *equant* was a point from which the center of an epicycle appeared to move at a constant rate. That meant the speed of the planets would vary slightly as viewed from Earth.

Ptolemy lived about five centuries after Aristotle, and although Ptolemy based his work on the Aristotelian universe, he was interested in a specific problem—the motions of the planets. Ptolemy was a brilliant mathematician, and he was mainly interested in creating a mathematical description of the motions he saw in the heavens. For him, first principles took second place to mathematical precision.

Aristotle's universe, as embodied in the mathematics of Ptolemy, dominated ancient astronomy, but it was wrong. The planets do not follow circles at uniform speeds. At first, the Ptolemaic system predicted the positions of the planets well, but as centuries passed, errors accumulated. If your watch gains only one second each day, it will keep time pretty well for months, but the error will gradually become noticeable. So, too, the errors in the Ptolemaic system gradually accumulated as the centuries passed, but because of the deep respect people had for the writings of Aristotle, the Ptolemaic system was not abandoned.

Islamic and later European astronomers tried to update the system, computing new constants and adjusting epicycles. In the middle of the 13th century a team of astronomers supported by King Alfonso X of Castile studied the *Almagest* for 10 years. Although they did not revise the theory very much, they simplified the calculation of the positions of the planets using the Ptolemaic system and published the result as the *Alfonsine Tables,* named in honor of the king, the last great attempt to make the Ptolemaic system of practical use.

Practicing SCIENCE

How did the models of Hipparchus and Ptolemy violate the principles of the early Greek philosophers Plato and Aristotle? Today, scientific reasoning depends on evidence and theory, but in classical times scientific reasoning almost always started from first principles.

Hipparchus and Ptolemy lived very late in the history of classical astronomy, and they concentrated more on the mathematical problems and less on philosophical principles. They replaced the perfect spheres of Plato with interlocking circles in the form of epicycles and deferents. They moved Earth slightly away from the center of the deferent, so their models were not exactly geocentric. Moreover, the epicycles moved uniformly only as seen from the equants. Thus, celestial motions in the model were no longer precisely uniform.

The achievements of Hipparchus and Ptolemy were steps on the way to the modern way of practicing science, in which scientists work to question all of their own and each other's assumptions. Scientists try not to consider any principles as being unquestionable.

4-2 The Copernican Revolution

You would not have expected Nicolaus Copernicus (1473–1543) to trigger a revolution in astronomy and science. He was born to a prosperous and politically connected merchant family in Poland. Orphaned at the age of 10, he was raised by his uncle, an important bishop, who sent him to the University of Krakow and then to the best universities in Italy. Copernicus studied law and medicine before pursuing a lifelong career as a Church administrator. Nevertheless, he had a passion for astronomy (**Figure 4-7**).

4-2a Copernicus and the Heliocentric Hypothesis

If you could go back in time and sit beside Copernicus in his astronomy classes, you would study the Ptolemaic model, a detailed version of Aristotle's universe. The central location of

Iryna1/Shutterstock.com

▲ **Figure 4-7** Nicolaus Copernicus (Latinized version of his birth name, Mikolaj Kopernik) pursued a lifetime career in the Church, but he was also a talented mathematician and astronomer. His work triggered a revolution in human thought.

Earth was widely accepted, and everyone knew that the heavens moved in uniform circular motion. For most scholars, questioning these principles was not an option because, over the course of centuries, Aristotle's universe had become linked with Christian theology. According to the Aristotelian view, the most perfect region was in the heavens and the most imperfect at Earth's center. That classical geocentric universe matched the commonly held Christian geometry of heaven and hell, so anyone who criticized the Ptolemaic model was not only questioning Aristotle's universe but also indirectly challenging belief in heaven and hell.

For this reason, Copernicus probably found it difficult at first to consider alternatives to the Ptolemaic universe. Throughout his life, he was associated with the Catholic Church, which had adopted many of Aristotle's ideas. Through the influence of his uncle the bishop, Copernicus was appointed a canon (a type of Church official) of the cathedral in Frauenberg at the unusually young age of 24. This gave Copernicus an income, although he continued his studies at the universities in Italy. When he finally left Italy, he joined his uncle and served as his secretary and personal physician until the uncle's death in 1512. At that point, Copernicus moved into quarters adjoining the cathedral in Frauenberg, where he lived and worked for the rest of his life.

His close connection with the Church notwithstanding, Copernicus began to consider an alternative to the Ptolemaic model, probably while he was still at university. Sometime before 1514, he wrote an essay proposing a heliocentric universe model in which the Sun, not Earth, is the center of the Universe. To explain the daily and annual cycles of the sky, Copernicus proposed that Earth rotates on its axis and revolves around the Sun. He distributed this commentary in handwritten form, without a title, and in some cases anonymously, to friends and astronomical correspondents. He may have been cautious out of modesty, out of respect for the Church, or out of fear that his unorthodox ideas would be attacked unfairly. Although his early essay discussed every major aspect of his later work, it did not include observations and calculations. His ideas needed supporting evidence, so Copernicus began gathering observations and making detailed calculations to be published as a book that would demonstrate the plausibility of his revolutionary idea.

4-2b Copernicus's Book: *De Revolutionibus*

Copernicus worked on his book *De Revolutionibus Orbium Coelestium (On the Revolutions of the Celestial Spheres)* over a period of many years and was essentially finished by about 1529. Some astronomers already knew of his work, and even some Church officials sought his advice on the reform of the calendar and looked forward to the book's publication. Nevertheless, Copernicus hesitated to publish it.

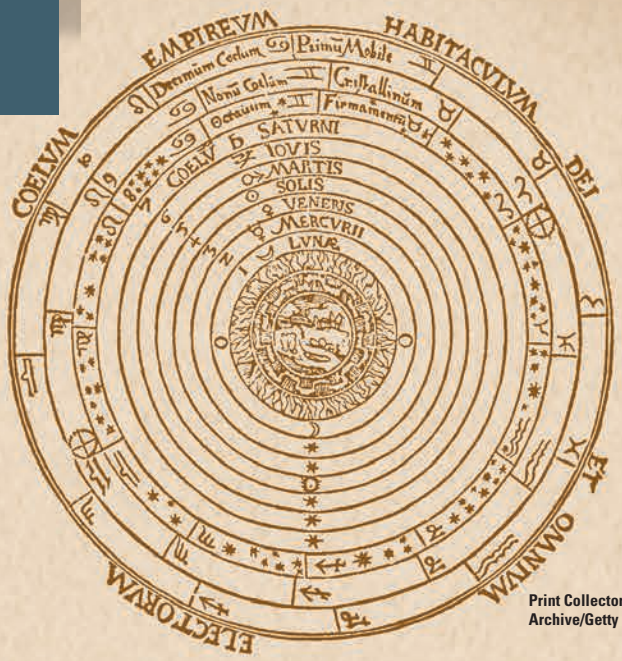
One reason he hesitated was that he knew the idea of a heliocentric universe would be highly controversial. This was a time of rebellion in the Church; Martin Luther was criticizing many Church teachings, and others—both scholars and scoundrels—were questioning the Church's authority. As you have learned, Earth's place in the Universe was linked in people's minds to the locations of heaven and hell, so moving Earth from its central place could be considered a heretical idea. Even matters as abstract as astronomy had a potential to stir possibly dangerous controversy.

Another reason Copernicus may have hesitated to publish his work was that it was incomplete. His model could not accurately predict planetary positions, so he continued to refine it. Finally in 1540 he allowed the visiting astronomer Joachim Rheticus to publish an account of the Copernican universe in Rheticus's book *Narratio Prima (First Narrative).* In late 1542, Copernicus sent the manuscript of *De Revolutionibus* off to be printed. He died in the spring of 1543 before the printing was completed.

The most important idea in the book was placing the Sun at the center of the Universe. That single innovation had an astonishing consequence: The retrograde motion of the planets was immediately explained in a straightforward way without the many epicycles used by Ptolemy.

In the Copernican system, Earth moves faster along its orbit than the planets that lie farther from the Sun. Consequently, Earth periodically overtakes and passes these planets. To visualize this, imagine that you are in a race car, driving rapidly along the inside lane of a circular racetrack. As you pass slower cars driving in the outer lanes, they fall behind, and if you did not realize you were moving, it would look as if the cars in the outer lanes occasionally slowed to a stop and then backed up for a short interval

An Ancient Model of the Universe

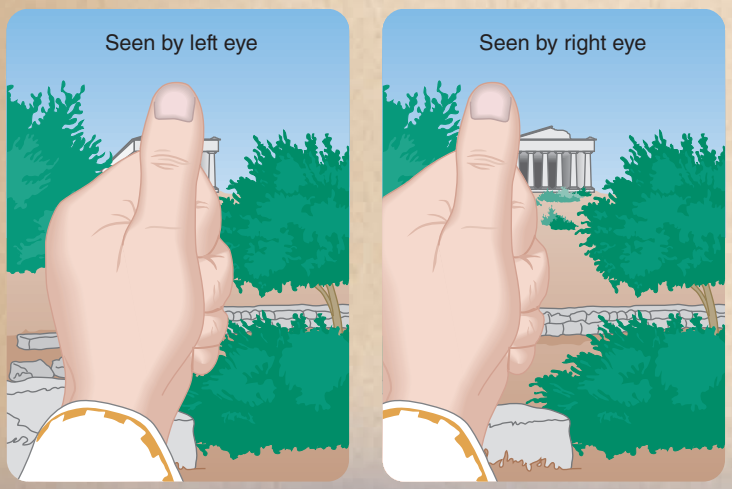


Print Collector/Hulton Archive/Getty Images

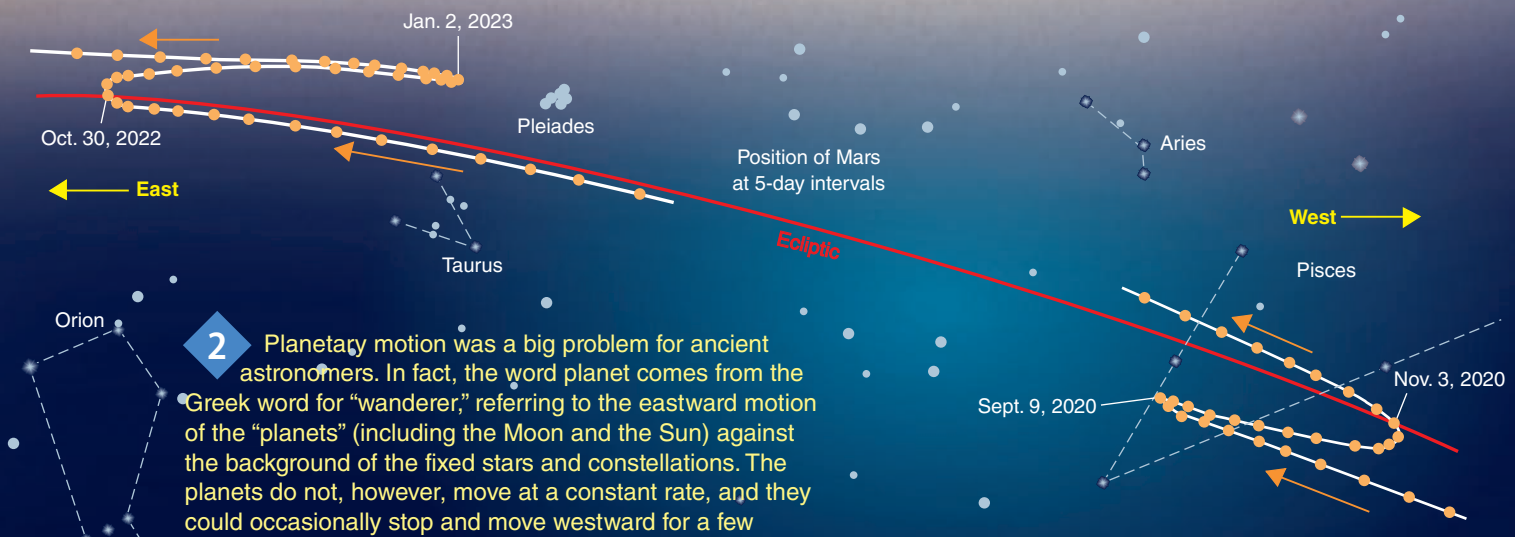
1 For 2000 years, the minds of astronomers were shackled by a pair of ideas. The Greek philosopher Plato argued that the heavens were perfect. Because he considered the only perfect geometrical shape to be a sphere, and a turning sphere carries a point on its surface around in a circle, and because the only perfect motion is uniform motion, Plato concluded that all motion in the heavens must be made up of combinations of circles turning at uniform rates. This idea was called *uniform circular motion*.

Plato's student Aristotle argued that Earth was imperfect and lay at the center of the Universe. Such a model is known as a *geocentric universe*. His model contained 55 spheres turning at different rates and at different angles to carry across the sky the seven celestial objects called planets at the time (the Moon, Mercury, Venus, the Sun, Mars, Jupiter, and Saturn).

Aristotle was known as the greatest philosopher in the ancient world, and for 2000 years his authority limited the imaginations of astronomers to uniform circular motion and geocentrism. See the model at right.



1a Early astronomers believed that Earth did not move because they saw no **parallax**, the apparent motion of an object because of the motion of the observer. (To demonstrate parallax, close one eye and cover a distant object with your thumb held at arm's length. Switch eyes, and your thumb seems to shift position, as shown at left.) If Earth moves, they reasoned, you should see the sky from different locations at different times of the year, and you should see parallax distorting the shapes of the constellations. They saw no parallax, so they concluded Earth does not move. Actually, the parallax of the stars is too small to see with the unaided eye.



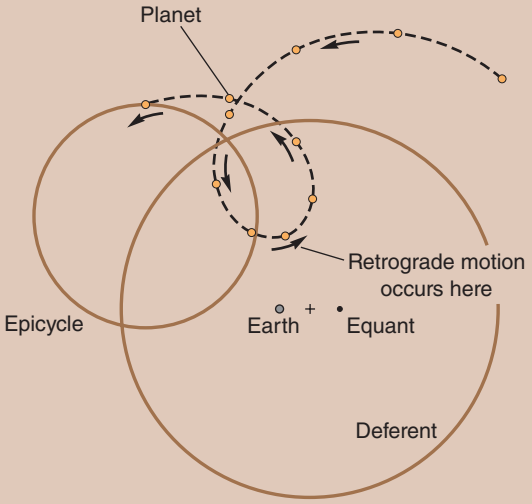
2 Planetary motion was a big problem for ancient astronomers. In fact, the word planet comes from the Greek word for "wanderer," referring to the eastward motion of the "planets" (including the Moon and the Sun) against the background of the fixed stars and constellations. The planets do not, however, move at a constant rate, and they could occasionally stop and move westward for a few months before resuming their eastward motion. This backward motion is called **retrograde motion**.

2a Every 2.14 years, Mars passes through a retrograde loop. Two successive loops are shown here. Each loop occurs farther east along the ecliptic and has its own shape. Simple uniform circular motion centered on Earth could not

3 Uniformly rotating circles were key elements of ancient astronomy. Ptolemy created a mathematical model of the Aristotelian universe in which the planet followed a small circle called an **epicycle** that slid around a larger circle called a **deferent**. By adjusting the size and rate of rotation of the circles, he could approximate the retrograde motion of a planet. See illustration at right.

To adjust the speed of the planet, Ptolemy supposed that Earth was slightly off-center, and that the center of the epicycle moved such that it appeared to move at a constant rate as seen from a point called the **equant**.

To further adjust his model, Ptolemy added small epicycles (not shown here) riding on top of larger epicycles, producing a highly complex model.

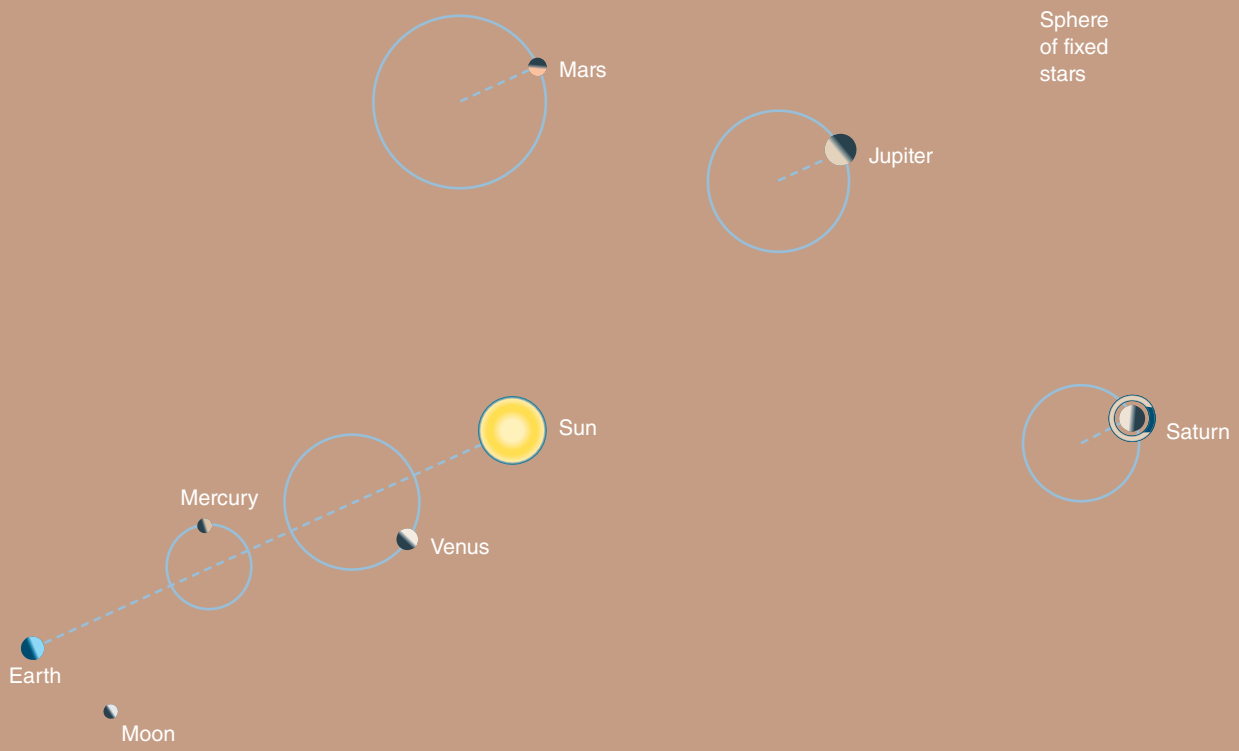


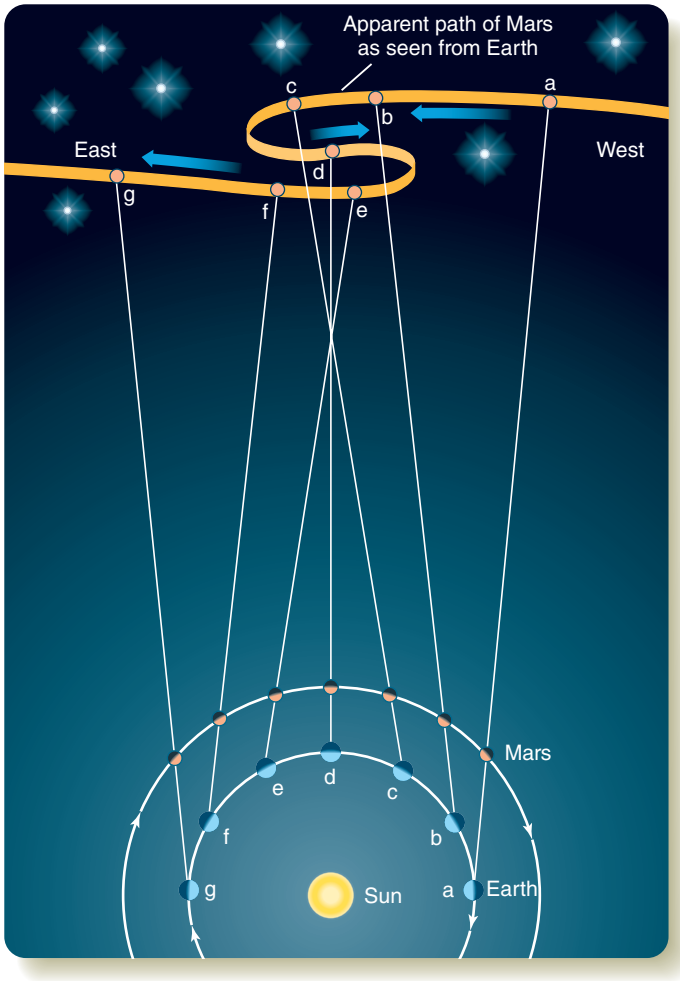
3a Ptolemy's great book *Mathematike Syntaxis,* published in Greek around the year 140, contained the details of his model. Islamic astronomers preserved and studied the book through the Middle Ages; they called it *Al Magisti (The Greatest)*. When the book was translated from Arabic to Latin in the 12th century, it became known as the *Almagest*.

3b The Ptolemaic model of the Universe shown below was geocentric and based on uniform circular motion. Note that Mercury and Venus were treated differently from the rest of the planets. The centers of the epicycles of Mercury and Venus had to remain on the Earth–Sun line as the Sun circled Earth through the year.

Equants and smaller epicycles are not shown here. Some versions of the model contained nearly 100 epicycles added by generations of astronomers trying to fine-tune the model to better reproduce the observed motion of the planets.

Notice that this modern illustration shows rings around Saturn and sunlight illuminating the globes of the planets, features that would not have been known before the invention of the telescope.





▲ **Figure 4-8** The Copernican explanation of retrograde motion. As Earth overtakes Mars (positions a through c), Mars appears to slow its forward motion. As Earth passes Mars (d), Mars appears to move backward. As Earth draws ahead of Mars (e–g), Mars resumes its forward motion against the background stars. The positions of Earth and Mars are shown at equal intervals of one month. (Compare with the Mars retrograde loop star charts on the left page of the Concept Art: An Ancient Model of the Universe.)

as you lapped them. **Figure 4-8** shows how the same thing happens as Earth passes a planet such as Mars. Although Mars moves steadily along its orbit, as seen from Earth it appears to slow to a stop and move westward (backward, retrograde) as Earth passes it. This happens to any planet whose orbit lies outside Earth's orbit, so Mars, Jupiter, and Saturn are all seen occasionally to move retrograde along the ecliptic. Because the planetary orbits do not lie precisely in the ecliptic, a planet does not resume its eastward, forward motion in precisely the same path it followed before the retrograde motion began. Consequently, the planet's path describes a loop with a shape that depends on the angle between the planet's orbital plane and the ecliptic and on the planet's location along the ecliptic.

Copernicus could explain retrograde motion without epicycles, and that was impressive. The Copernican system was elegant and simple compared with the multiple whirling epicycles and off-center equants of the Ptolemaic model. You can see Copernicus's own diagram for his heliocentric model in **Figure 4-9a**. Another advantage of the Copernican model was that in the Ptolemaic model, Mercury and Venus had to be treated differently from the rest of the planets; their epicycles had to remain centered on the Earth–Sun line because they are never seen far from the Sun. In contrast, in the Copernican model, all of the planets were treated the same. They all followed orbits that circled the Sun at the center. Furthermore, their speed depended in an orderly way on their distance from the Sun, with those closest moving fastest (**Figure 4-9b**).

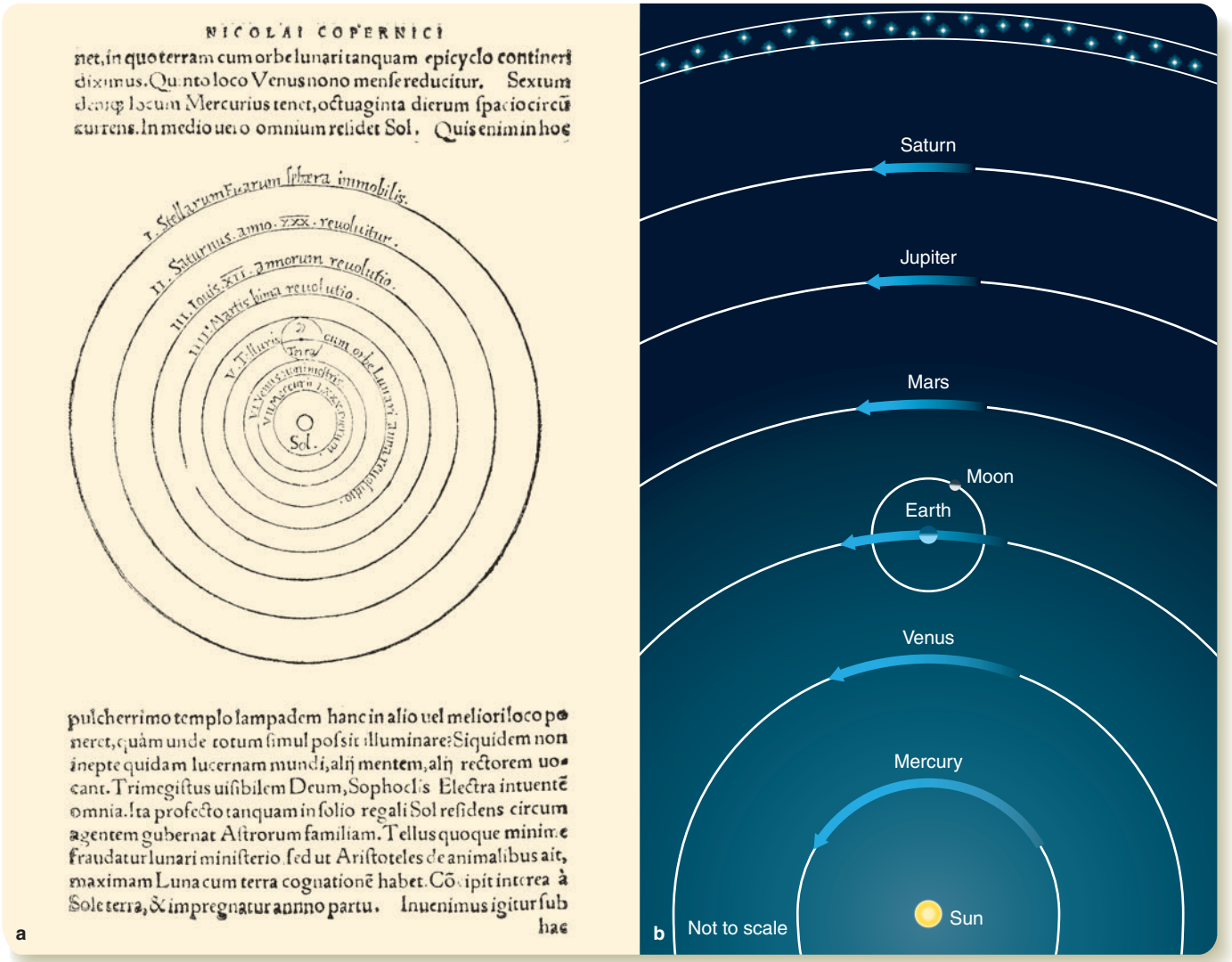
It is important to note that *De Revolutionibus* failed in one critical way: The Copernican model could not predict the positions of the planets any more accurately than the Ptolemaic model could. To understand why it failed, you need to understand Copernicus's mind-set.

Copernicus proposed an astonishingly bold idea when he made the Universe (meaning, the Solar System) heliocentric. Nevertheless, Copernicus was a classically trained astronomer with tremendous respect for the old concept of uniform circular motion. In fact, Copernicus objected strongly to Ptolemy's use of the equant. It seemed arbitrary to Copernicus, an obvious violation of the elegance of Aristotle's philosophy of the heavens. Copernicus called equants "monstrous" because they undermined both geocentrism and uniform circular motion. In devising his model, Copernicus managed to abandon geocentrism but not his strong belief in uniform circular motion.

Although he did not need epicycles to explain retrograde motion, Copernicus quickly discovered that the Sun, Moon, and planets showed other small variations in their motions that he could not explain using uniform circular motion centered on the Sun. Today astronomers recognize those variations as the result of planets following orbits that are not quite circular; the orbits actually are very slightly elliptical. Because Copernicus held firmly to uniform circular motion, he had to include his own new, small epicycles to try to reproduce these minor variations observed in the motions of the objects in the Solar System.

Because Copernicus retained uniform circular motion in his model, it could not accurately predict the motions of the planets. Only 9 years after the publication of the Copernican theory, a set of planetary tables called the *Prutenic Tables* (named after Prussia) was calculated and published by the astronomer Erasums Reinhold. Although these tables were based on the new Copernican model, they were not significantly more accurate than the 300-year-old *Alfonsine Tables* that were based on Ptolemy's model. Both could be in error by as much as 2 degrees, which is four times the angular diameter of the full moon.

At this point you should consider an important distinction: The Copernican *model* is inaccurate. It includes uniform circular motion and consequently does not precisely describe the motions of the planets. But the Copernican *hypothesis*—that the Solar System is heliocentric—is correct. The planets in fact circle the Sun, not Earth.


NICOLAI COPERNICI
net, in quo terram cum orbe lunari tanquam epicyclo contineri
diximus. Quinto loco Venus nono mense reducitur. Sextum
denique locum Mercurius tenet, octuaginta dierum spacio circũ
currens. In medio uero omnium relidet Sol. Quis enim in hoc
I. Stellarum fixarum sphæra immobilis.
II. Saturnus. anno. XXX. revolvitur.
III. Iouis. XII. annorum revolutio.
IIII. Martis bima revolutio.
V. Tellus. cum orbe lunari. anua revolutio.
VI. Veneris nouemestris
VII. Mercurius LXXX
Terra
Sol.
pulcherrimo templo lampadem hanc in alio uel meliori loco po
neret, quàm unde totum simul possit illuminare? Siquidem non
inepte quidam lucernam mundi, alij mentem, alij rectorem uo
cant. Trimegistus uisibilem Deum, Sophoclis Electra intuentẽ
omnia. Ita profecto tanquam in solio regali Sol residens circum
agentem gubernat Astrorum familiam. Tellus quoque minime
fraudatur lunari ministerio sed ut Aristoteles de animalibus ait,
maximam Luna cum terra cognationẽ habet. Cõcipit interea à
Sole terra, & impregnatur anuno partu. Inuenimus igitur sub
hac
a
Saturn
Jupiter
Mars
Moon
Earth
Venus
Mercury
Sun
Not to scale
b

From the collection of Yerkes Observatory

▲ **Figure 4-9** (a) The Copernican universe as drawn by Copernicus in his book *De Revolutionibus*. Earth and all the known planets revolve in separate circular orbits about the Sun (Sol) at the center. The outermost sphere carries the immobile stars of the celestial sphere. Notice the orbit of the Moon around Earth (Terra). (b) The model was simple not only in its arrangement of the planets but also in the explanation of their motions. Orbital speed (blue arrows) decreased from Mercury, the fastest, to Saturn, the slowest. Compare the elegance of this model with the complexity of the Ptolemaic model in Figure 3a of the Concept Art.

Although astronomers throughout Europe read and admired *De Revolutionibus,* they did not immediately accept the Copernican hypothesis. The mathematics were elegant, and the astronomical observations and calculations were of tremendous value, but few astronomers believed, at first, that the Sun actually was the center of the planetary system and that Earth moved. How the Copernican hypothesis was gradually recognized as correct has been called the Copernican Revolution because it was not just the adoption of a new idea but a total change in the way astronomers, and, truly, all of humanity, thought about the place of Earth. In fact, our modern use of the words *revolution* and *revolutionary* to describe philosophical, political, and social upheavals comes from the title of Copernicus's book, *On Revolutions* (**How Do We Know? 4-1**).

There are several reasons the Copernican hypothesis gradually won support, but the most important factor may have been the beautiful simplicity of the idea. Placing the Sun at the center of the Universe produced a symmetry among the motions of the planets that is pleasing to the eye as well as to the intellect.

The most astonishing consequence of the Copernican hypothesis was not what it said about the Sun but what it said about Earth. Placing the Sun at the center, Copernicus made Earth into a planet moving along an orbit like the other planets. By making Earth a planet, Copernicus completely changed humanity's view of our place in the Universe and triggered a controversy that would eventually bring the astronomer Galileo Galilei before the Inquisition. This controversy over the apparent conflict between scientific knowledge and philosophical and theological ideas continues even today.

Scientific Revolutions

How do scientific revolutions occur?

You might think from what you know of the scientific method that science grinds forward steadily as new hypotheses are tested against evidence and accepted or rejected. In fact, science sometimes leaps forward in scientific revolutions. The Copernican Revolution is often cited as the perfect (and first) example. In the course of a few decades, astronomers rejected the 2000-year-old geocentric model and adopted the heliocentric model. Why does that happen? It's all because scientists are human.

The philosopher of science Thomas Kuhn has referred to a commonly accepted set of scientific ideas and assumptions as a scientific **paradigm** (pronounced *PEAR-ah-dime*). The pre-Copernican astronomers shared a geocentric paradigm that included uniform circular motion, geocentrism, and the perfection of the heavens. Although they were intelligent, they were prisoners of that paradigm. A scientific paradigm is powerful because it shapes your perceptions. It determines what you judge to be important questions and what you judge to be significant evidence. Consequently, the ancient astronomers could not recognize how their geocentric paradigm limited their understanding.

The works of Copernicus, Galileo, and Kepler overthrew the geocentric paradigm. Scientific revolutions occur when the deficiencies of the old paradigm build up until finally someone has the insight to think "outside the box." Pointing out the failings of the old ideas and proposing a new paradigm with supporting evidence is like poking a hole in a dam; suddenly the pressure is released, and the old paradigm is swept away.

Scientific revolutions are exciting because they give you a dramatic new understanding of nature, but they are also times of conflict as new insights sweep away old ideas.

NYPL/Science Source/Getty Images

People originally believed stars are attached to the surface of a sphere that enclosed the Universe.

Practicing SCIENCE

Why would you say the Copernican hypothesis was correct but the Copernican model was inaccurate?

Distinguishing between hypotheses and models is an important part of practicing science.

The Copernican hypothesis was that the Sun and not Earth is the center of the Universe. Given the limited knowledge of the Renaissance astronomers about distant stars and galaxies, that hypothesis was correct.

The Copernican model, however, included not only the heliocentric hypothesis but also an assumption of uniform circular motion. The model is inaccurate because the planets don't really follow circular orbits, nor travel at constant rates, and the small epicycles that Copernicus added to his model never quite reproduced the motions of the planets.

Now, look back to Section 2-2a. The celestial sphere is a model. ***How is it accurate—how does it allow precise calculations of astronomical phenomena? If, instead, the celestial sphere were to be considered a hypothesis, is it correct?***

4-3 Tycho, Kepler, and Planetary Motion

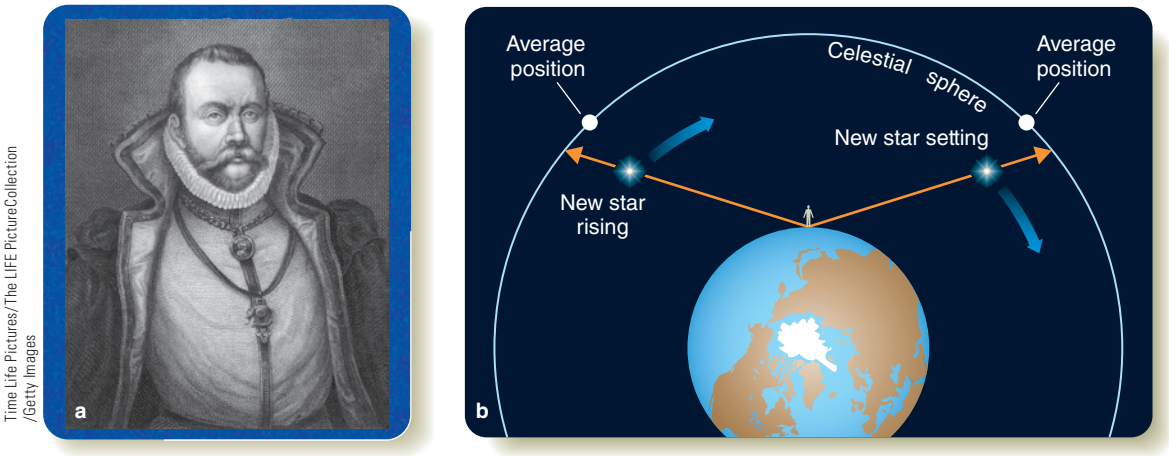
The Copernican hypothesis solved the problem of the place of Earth, but it did not completely explain the details of planetary motion. If planets do not move in uniform circular motion, how do they move? The puzzle of planetary motion was solved during the century following the death of Copernicus almost entirely through the work of two scientists. One compiled the observations, and the other did the analysis.

4-3a Tycho Brahe

Tycho Brahe (1546–1601), known today simply as Tycho (pronounced *TEE-co*), was not a churchman like Copernicus but rather a nobleman from an important family, but like Copernicus he was educated at the finest universities. He was infamous for his vanity and his lordly, haughty manners. Tycho's disposition perhaps was not improved by a dueling injury from his university days. His nose was badly disfigured, and for the rest of his life he wore false noses made of gold and silver, stuck on with wax.

Although Tycho officially studied law at the university, his real passions, much to his family's disappointment, were mathematics and astronomy, and early in his university days he began measuring the positions of the planets in the sky. In 1563, Jupiter and Saturn passed very near each other in the sky, nearly merging into a single point on the night of August 24. Tycho found that the old *Alfonsine Tables,* based on the Ptolemaic model, were a full month in error in their prediction of that event, and the new *Prutenic Tables,* based on the Copernican, model were in error by several days.

In 1572, a "new star" (now called Tycho's supernova) appeared in the sky, shining more brightly than Venus, and



▲ **Figure 4-10** (a) Tycho Brahe was, during his lifetime, the most famous astronomer in the world. (b) In Aristotle's universe, a "new star" must be located below the sphere of the Moon and not on the celestial sphere because it would be temporary, not eternal. Consequently, reasoned Tycho, the new star of 1572 should display parallax and be seen east of its average position as it rose, west of its average position as it set. Because he did not detect any daily parallax—the new star stayed put exactly at its average position—Tycho concluded that the new star had to lie on the celestial sphere, far beyond the Moon.

Tycho carefully measured its position. According to classical astronomy, the new star represented a change in the heavens and therefore had to lie below (closer than) the sphere of the Moon. To Tycho, who at this time still believed in a geocentric universe, this meant that the new star should show parallax, meaning that it would appear slightly too far east as it rose and slightly too far west as it set (**Figure 4-10**). But Tycho saw no parallax in the position of the new star, so he concluded that it must lie above the sphere of the Moon and was probably on the starry sphere itself. This contradicted Aristotle's conception of the starry sphere as perfect and unchanging.

No one before Tycho could have made this discovery because no one had ever measured the positions of celestial objects as accurately as he did. Tycho had great confidence in the precision of his measurements, and he had studied astronomy thoroughly, so when he failed to detect parallax for the new star, he knew it was important evidence against the Ptolemaic theory. He announced his discovery in a small book, *De Stella Nova (On the New Star),* published in 1573.

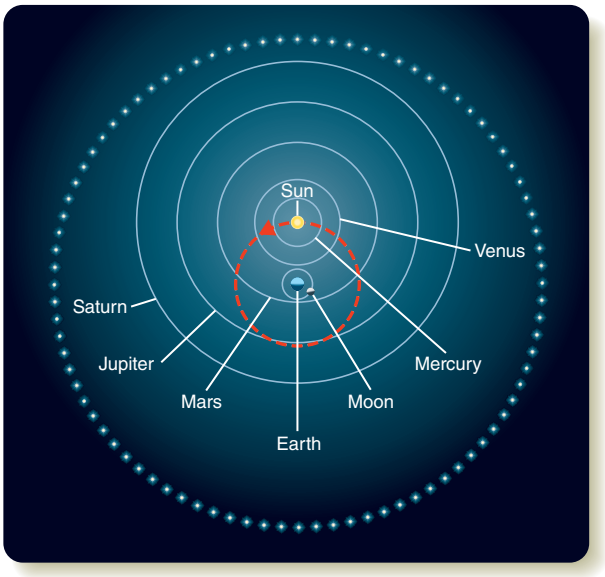
The book attracted the attention of astronomers throughout Europe, and soon Tycho's family, concerned about his professional future, introduced him to the court of the Danish King Frederick II, where he was offered funds to build an observatory on the island of Hveen just off the Danish coast. To support his observatory, Tycho was given a steady income as lord of a coastal district from which he collected rents. (It is said that he was not a popular landlord.) On Hveen, Tycho constructed a luxurious home with six towers especially equipped for astronomy and populated it with servants, assistants, and a dwarf to act as jester. Soon Hveen was an international center of astronomical study.

4-3b Tycho's Legacy

Tycho made no direct contribution to astronomical theory. Because he could measure no parallax for the stars, he concluded that Earth had to be stationary, thus rejecting the Copernican hypothesis. However, he also rejected the Ptolemaic model because of its inaccuracy. Instead, he devised a complex model in which Earth was the unmoving center of the Universe around which the Sun and the Moon moved while the other planets circled the Sun (**Figure 4-11**). The model thus incorporated part of the Copernican model, but Tycho preserved the central immobile Earth. Although Tycho's model was popular at first, the Copernican model replaced it within a century.

The true value of Tycho's work was the quality of his observational data. Because he was able to devise new and better instruments, he was able to make highly accurate observations of the position of the stars, Sun, Moon, and planets. Tycho had no telescopes—they were not invented until the following century—so his observations were made by naked eye, peering along devices much like gun sights. He and his assistants made precise observations for 20 years at Hveen.

Unhappily for Tycho, King Frederick II died in 1588, and his young son took the throne. Suddenly, Tycho's temper, vanity, and noble presumptions removed him from favor. In 1596, taking most of his instruments and books of observations, he went to Prague, the capital of Bohemia, and became imperial mathematician to the Holy Roman Emperor Rudolph II. His goal was to revise the *Alfonsine Tables* and publish the result as a monument to his new patron. He promised to call it the *Rudolphine Tables.*



▲ **Figure 4-11** Tycho's model of the Universe retained the first principles of classical astronomy; it was geocentric with the Sun and Moon revolving around Earth, but the planets revolved around the Sun. All motion was along circular paths.

Tycho did not intend to base the *Rudolphine Tables* on the Ptolemaic system but rather on his own Tychonic system, proving once and for all the validity of his hypothesis. To assist him, he hired a few mathematicians and astronomers, including one Johannes Kepler. Then, in November 1601, Tycho collapsed while visiting a nobleman's home in Prague. Before he died 11 days later, he asked Rudolph II to make Kepler imperial mathematician. The newcomer, not a nobleman at all but a commoner, became Tycho's replacement (at one-sixth Tycho's salary).

4-3c Johannes Kepler

No one could have been more different from Tycho Brahe than Johannes Kepler (**Figure 4-12a**). Kepler was born in 1571 to a poor family in a region that is now part of southwest Germany. His father was unreliable and shiftless, principally employed as a mercenary soldier fighting for whoever paid enough. He was often absent for long periods and finally failed to return from a military expedition. Kepler's mother was apparently an unpleasant and unpopular woman. She was accused of witchcraft, and Kepler had to defend her in a trial that dragged on for 3 years. She was finally acquitted but died the following year.

Despite family disadvantages and chronic poor health, Kepler did well in school, winning promotion to a Latin school and eventually a scholarship to the university at Tübingen, where he studied to become a Lutheran pastor. During his last year of study, Kepler accepted a job in Graz teaching mathematics and astronomy. His superiors put him to work teaching a few introductory courses and preparing an annual almanac that contained astronomical, astrological, and weather predictions. Evidently he was not a good teacher; he had few students his first year and none at all his second. Fortunately, and of course only by pure luck, some of his predictions were considered to be fulfilled, and he gained a reputation as an astrologer and seer. Even in later life he earned money from his almanacs.

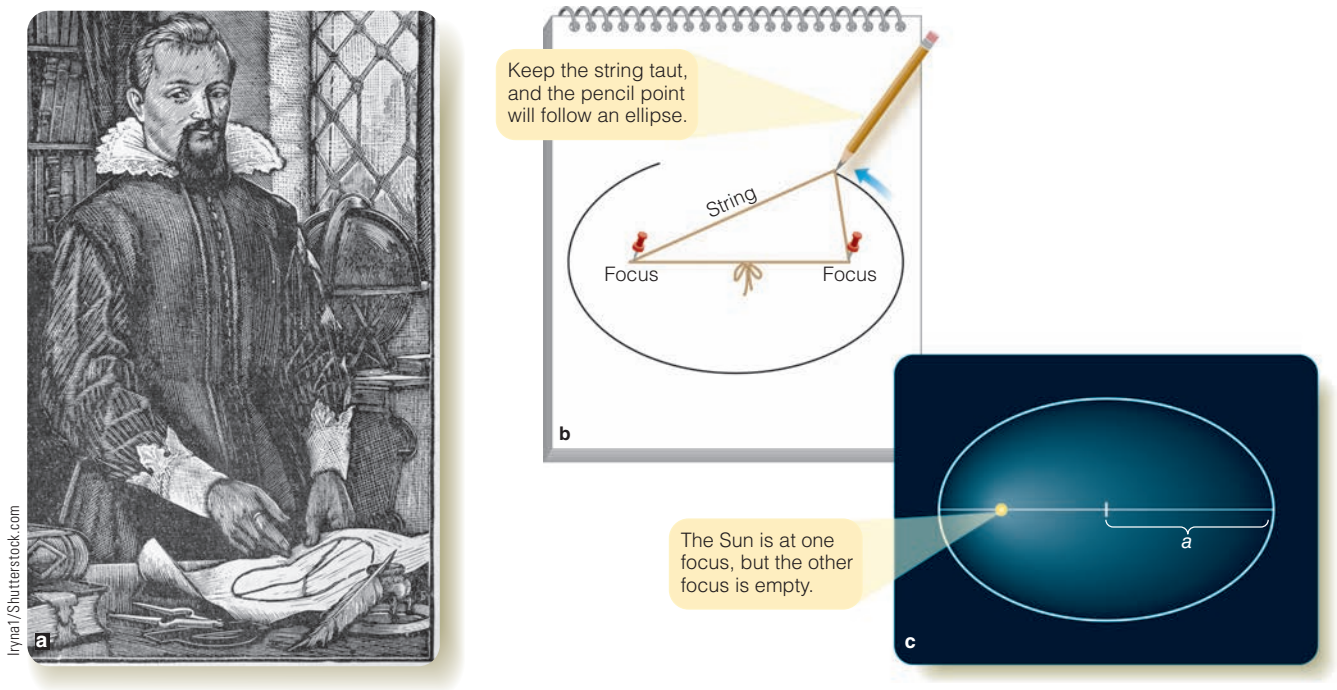
While still a college student, Kepler had become a believer in the Copernican hypothesis, and at Graz he used his extensive spare time to study astronomy. By 1596, the same year Tycho arrived in Prague, Kepler was sure he had solved the mystery of the Universe. That year he published a book in Latin with a 28-word title that is usually abbreviated as *Mysterium Cosmographicum (Mystery of the Universe)*.

By modern standards, *Mysterium Cosmographicum* contains almost nothing of value. It begins with a long appreciation of Copernicanism and then goes on to speculate on the reasons for the spacing of the planetary orbits. Kepler believed that he had found the underlying architecture of the Universe in the properties of spheres plus five special shapes known as the *regular solids*, which include cubes and tetrahedrons. In Kepler's model, the five regular solids became spacers for the orbits of the six planets. Kepler went so far as to conclude that there must be only six planets (Mercury, Venus, Earth, Mars, Jupiter, and Saturn) because there were only five regular solids to act as spacers between their spheres. This hypothesis is considered nonsense today, but *Mysterium Cosmographicum* had one virtue: As Kepler tried to fit the five solids to the planetary orbits, he demonstrated that he was a talented mathematician and that he was well versed in astronomy. He sent copies of his book to well-known astronomers, including Tycho and Galileo.

4-3d Working with Tycho

Life became unsettled for Kepler because of recurring persecution of Protestants in the region where he lived, so when Tycho Brahe invited him to Prague in 1600, Kepler went readily, eager to work with a famous astronomer. Tycho's sudden death in 1601 left Kepler, the new imperial mathematician, in a position to use Tycho's data to analyze the motions of the planets and complete the *Rudolphine Tables.* Tycho's family, recognizing that Kepler was a Copernican and guessing that he would not follow the Tychonic system in completing the *Rudolphine Tables,* sued to recover the instruments and books of observations. The legal wrangle went on for years. Tycho's family did get back the instruments Tycho had brought to Prague, perhaps because Kepler, with his poor eyesight, couldn't use them, but Kepler had the books of observational data, and he kept them.

Whether Kepler had any legal right to Tycho's records is debatable, but he put them to good use. He began by studying the motion of Mars, trying to deduce from the observations how the planet moved. By 1606, he had solved the mystery. The orbit of Mars is not a circle but an ellipse. With that, Kepler gave up the 2000-year-old belief in the circular motion of the planets. But even this insight was not enough to explain the observations. In fact, the planets do not move at uniform speeds along their elliptical orbits.



▲ **Figure 4-12** (a) Johannes Kepler was Tycho Brahe's successor. (b) The geometry of elliptical orbits: Drawing an ellipse with two tacks and a loop of string is easy. (c) The semimajor axis, *a*, is half of the longest diameter. The Sun lies at one of the foci of the elliptical orbit of a planet.

Kepler's analysis showed that they move faster when close to the Sun and slower when farther away. With those two brilliant discoveries, Kepler abandoned both uniform motion and circular motion, and thereby finally solved the puzzle of planetary motion. He published his results in 1609 in a book called *Astronomia Nova (New Astronomy).*

Despite the abdication of Rudolph II in 1611, Kepler continued his astronomical work. He wrote about a supernova that appeared in 1604 (now known as Kepler's supernova) and about comets, and he produced a textbook about Copernican astronomy. In 1619, he published *Harmonices Mundi (The Harmony of the World),* in which he returned to the cosmic mysteries of *Mysterium Cosmographicum.* The only thing of note in *Harmonices Mundi* is Kepler's discovery that the radii of the planetary orbits are related to the planets' orbital periods. That and his two previous discoveries are so important that they have become known as the three fundamental laws of orbital motion.

4-3e Kepler's Three Laws of Planetary Motion

Although Kepler dabbled in the philosophical arguments of his day, he was at heart a mathematician, and his triumph was his mathematical explanation of the motion of the planets. The key to his solution was the ellipse.

An **ellipse** is a figure that can be drawn around two points, called the foci, in such a way that the distance from one focus to any point on the ellipse and back to the other focus equals a constant. This makes it easy to draw ellipses using two thumbtacks and a loop of string. Press the thumbtacks into a board, loop the string about the tacks, and place a pencil in the loop. If you keep the string taut as you move the pencil, it traces out an ellipse (**Figure 4-12b**).

The geometry of an ellipse is described by two simple numbers: (1) The **semimajor axis**, ***a***, is half of the longest diameter, as you can see in **Figure 4-12c**. (2) The **eccentricity**, ***e***, of an ellipse is half the distance between the foci divided by the semimajor axis. The eccentricity of an ellipse tells you its shape; if *e* is nearly equal to one, the ellipse is very elongated. If *e* is close to zero, the ellipse is more circular. To draw a circle with the string and tacks shown in Figure 4-12b, you would have to move the two thumbtacks together because a circle is really just an ellipse with eccentricity equal to zero.

Ellipses are an essential part of Kepler's three fundamental rules of planetary motion (**Table 4-1**). Those rules have been tested and confirmed so many times that astronomers now refer to them as natural laws.

Table 4-1 Kepler's Laws of Planetary Motion

I. The orbits of the planets are ellipses with the Sun at one focus.
II. A line from a planet to the Sun sweeps over equal areas in equal intervals of time.
III. A planet's orbital period squared is proportional to its average distance from the Sun cubed:

$$P_{yr}^2 = a_{AU}^3$$

Kepler's first law says that the orbits of the planets around the Sun are ellipses with the Sun at one focus. Thanks to the precision of Tycho's observations and the sophistication of Kepler's mathematics, Kepler was able to recognize the elliptical shape of the orbits even though they are nearly circular. Mercury has the most elliptical orbit, but even it deviates only slightly from a circle (**Figure 4-13a**).

Kepler's second law says that an imaginary line drawn from the planet to the Sun always sweeps over equal areas in equal intervals of time. This means that when the planet is closer to the Sun and the line connecting it to the Sun is shorter, the planet must move more rapidly so that the line sweeps over the same area per time interval that it sweeps over when the planet is farther from the Sun. For example, the hypothetical planet in **Figure 4-13b** that follows a highly elongated ellipse moves from point *A* to point *B* in one month, sweeping over the area shown. When the planet is farther from the Sun, one month's motion along the orbit would be less, from *A'* to *B'*, but the area swept out would be the same.

Kepler's third law relates a planet's orbital period to its average distance from the Sun. The orbital period, *P*, is the time a planet takes to travel around the Sun once. The average distance of a planet from the Sun around its elliptical path turns out simply to equal the semimajor axis of its orbit, *a*. Kepler's third law says that a planet's orbital period squared is proportional to the semimajor axis of its orbit cubed (**Figure 4-13c**). Measuring *P* in years and *a* in astronomical units, you can summarize the third law as:

KEY EQUATION 4-1 Kepler's Third Law

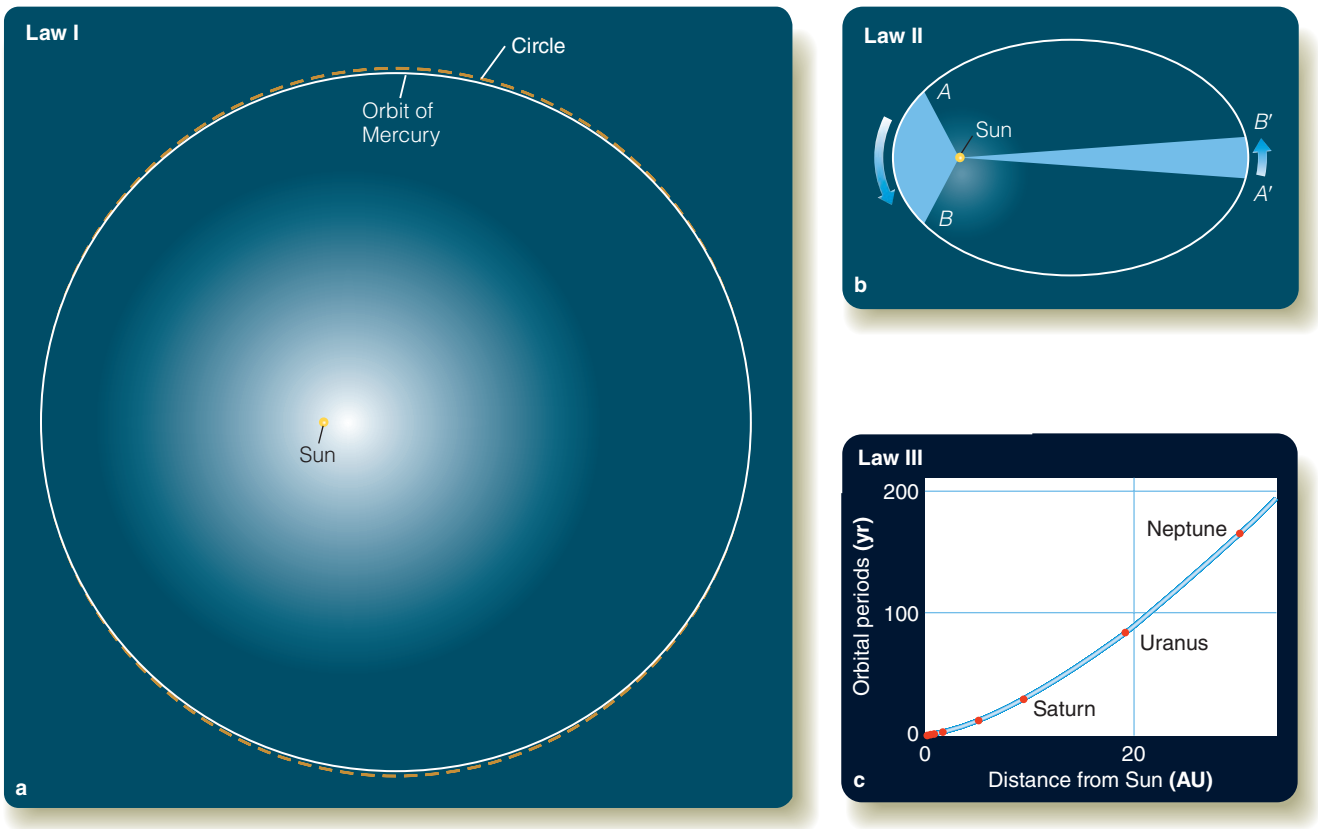
$$P_{\text{yr}}^2 = a_{\text{AU}}^3$$

Example: What is the orbital period of Jupiter? ***Solution:*** Jupiter's average distance from the Sun is 5.20 astronomical units (AU). That value cubed is about 141, so the period must be the square root of 141, which equals 11.9 years.

Notice that Kepler's three laws are **empirical**. That is, they describe a phenomenon without explaining why it occurs. Kepler derived the laws from Tycho's extensive observations, not from any first principle, fundamental assumption, or theory. In fact, Kepler never knew what held the planets in their orbits or why they continued to move around the Sun.

4-3f Kepler's Final Book: The *Rudolphine Tables*

Kepler continued his mathematical work on the *Rudolphine Tables,* and at last, in 1627, it was ready. He financed the printing himself, dedicating the book to the memory of Tycho Brahe. In



▲ **Figure 4-13** Kepler's three laws: (a) The first law says the orbits of the planets are not perfect circles but are slightly elliptical. In this scale drawing of the orbit of Mercury, it looks nearly circular. (b) The second law is demonstrated by a planet that moves from *A* to *B* in 1 month and also from *A'* to *B'* in 1 month. The two light blue sectors have the same area. (c) The third law defines how the orbital periods of the planets are related to their distance from the Sun.

Hypotheses, Theories, and Laws

Why is a theory much more than just a guess?

Scientists study nature by devising and testing new hypotheses and then developing the successful ideas into theories and laws that describe how nature works. A good example is the connection between sour milk and the spread of disease.

A scientist's first step in solving a natural mystery is to propose a reasonable explanation based on what is known so far. This proposal, called a **hypothesis**, is a single assertion or statement that can be tested through observations and experiments. If the explanation is not testable somehow, it is not really a scientific hypothesis.

From the time of Aristotle, philosophers believed that food spoils as a result of the spontaneous generation of life—for example, mold growing out of drying bread. French chemist Louis Pasteur hypothesized instead that microorganisms were not spontaneously generated but were carried through the air. To test his hypothesis, he sealed an uncontaminated nutrient broth in glass, completely protecting it from the spores, microorganisms, and dust particles in the air. No mold grew, effectively disproving spontaneous generation. Although others had argued against spontaneous generation before Pasteur, it was Pasteur's meticulous testing of his hypothesis through experimentation that finally convinced the scientific community.

A **theory** generalizes the specific results of well-confirmed hypotheses to give a broad description of nature that can be applied to a wide variety of circumstances. For instance, Pasteur's specific hypothesis about mold growing in broth contributed to a broader theory that disease is caused by microorganisms transmitted from sick people to healthy people. This theory, called the *germ theory of disease,* is a cornerstone of modern medicine. It is a **Common Misconception** that the word *theory* means a tentative idea, a guess. Actually, "hypothesis" is the word that scientists use to describe what a layperson would call a guess. Scientists generally use the word *theory* to mean an idea that is widely applicable, has been tested in many ways, and confirmed by abundant evidence—solid and trustworthy, much better than a guess.

Sometimes, when a theory has been refined, tested, and confirmed so often that scientists have great confidence in it, and is so basic that it is applicable everywhere and every time, it is called a **natural law**. Natural laws are the most fundamental principles of scientific knowledge. Kepler's laws of planetary motion are good examples.

Confidence is the key criterion. In general, scientists have more confidence in a theory than in a hypothesis and the most confidence in a natural law. However, there is no precise distinction among a hypothesis, a theory, and a law, and use of these terms is sometimes a matter of tradition. For instance, some textbooks refer to the Copernican "theory" of heliocentrism, but it had not been well tested when Copernicus proposed it, and it is more rightly called the Copernican hypothesis. At the other extreme, Darwin's "theory" of evolution, containing many hypotheses that have been tested and confirmed over and over for nearly 150 years, might more correctly be called a natural law.

Collection of John Coolidge III/Michael A. Seeds

A fossil of a 500-million-year-old trilobite is one piece of evidence confirming Darwin's theory of evolution, which is really a natural law.

fact, Tycho's name appears in larger type on the title page than Kepler's own. This is surprising because the tables were not based on the Tychonic system but on the heliocentric model of Copernicus and the elliptical orbits of Kepler. The reason for Kepler's evident deference was Tycho's family, still powerful and still intent on protecting Tycho's reputation. They even demanded a share of the profits and the right to censor the book before publication, though they changed nothing but a few words on the title page and added an elaborate dedication to the emperor.

The *Rudolphine Tables* was Kepler's masterpiece. The tables could predict the positions of the planets 10 to 100 times more accurately than previous tables. Kepler's tables were the precise model of planetary motion that Copernicus had sought but failed to find because he could not give up the idea of perfectly circular motions. The accuracy of the *Rudolphine Tables* was strong evidence that both Kepler's laws of planetary motion and the Copernican hypothesis for the place of Earth were correct. Copernicus would have been pleased.

Kepler died in 1630. He had solved the problem of planetary motion, and his *Rudolphine Tables* demonstrated his solution. Although he did not understand why the planets moved or why they followed ellipses—insights that had to wait half a century for Isaac Newton—Kepler's three laws worked. In science the only test of a hypothesis is, "Does it describe reality?" (In other words, does it match what is observed?) Kepler's laws have been used for almost four centuries as a true description of orbital motion. (**How Do We Know? 4-2**)

Practicing SCIENCE

How were Kepler's three laws of planetary motion based on evidence? Kepler derived his three laws of planetary motion from the observations made by Tycho Brahe during 20 years on Hveen.

The observations were the evidence, and they gave Kepler a reality check each time he tried a new calculation. He chose ellipses because they were the best fit to the data, not because he had some reason before he started his work to expect that the answer must be ellipses. Trying not to have a load of preconceptions about the final results, for example striving to identify and question first principles, is an important part of the modern way of practicing science.

The Copernican model was a poor predictor of planetary motion, but Kepler's *Rudolphine Tables* were much more accurate. ***Why might you expect that, given how the* Rudolphine Tables *were produced?***

4-4 Galileo's Conclusive Evidence

Most people think they know two "facts" about Galileo, but both facts are wrong They are **Common Misconceptions**, so you have probably heard them. The truth is, Galileo did not invent the telescope, nor was he condemned by the Inquisition for believing Earth moves around the Sun. Then why is Galileo so famous? Why did the Vatican reopen his case in 1979, almost 400 years after his trial? As you learn about Galileo, you will discover that his legacy concerns not only advancing understanding of the place of Earth and the motion of the planets but also helping establish a new and powerful method of understanding nature, a method called *science.* The scientific method requires combining discoveries by scientists like Galileo, who today would be called an experimentalist, with accomplishments of scientists like Kepler, who today would be called a theorist.

4-4a Telescope Observations

Galileo Galilei (1564–1642) (**Figure 4-14a**) was born in Pisa, a city in what is now Italy, and he studied medicine at the university there. His true love, however, was mathematics, and, although he had to leave school early for financial reasons, he returned only 4 years later as a professor of mathematics. Three years after that he became professor of mathematics at the university in Padua, where he remained for 18 years.

During this time, Galileo seems to have adopted the Copernican model, although he admitted in a 1597 letter to Kepler that he did not support Copernicanism publicly. At that time, the Copernican hypothesis was not officially considered heretical, but it was hotly debated among astronomers, and Galileo, living in a region controlled by the Catholic Church, cautiously avoided trouble. It was the telescope that finally drove Galileo to publicly defend the heliocentric model.

Iryna1/Shutterstock.com

Leemage/UIG/Getty Images

▲ **Figure 4-14** (a) Galileo is remembered as the great defender of Copernicanism. (b) Two of Galileo's telescopes, on display in a museum in Florence. Although he did not invent telescopes, Galileo will always be associated with them because they were the source of much of the observational evidence he used to try to understand the Universe.

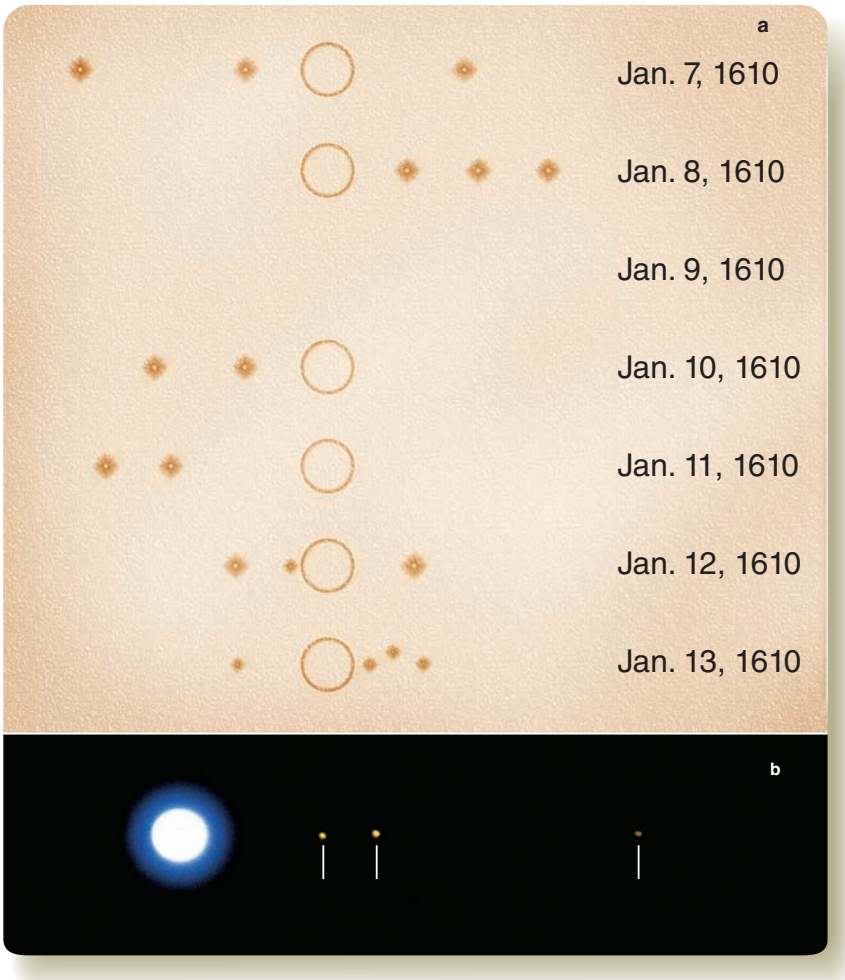
The telescope seems to have been invented around 1608 by lens makers in Holland. Galileo, hearing descriptions in the fall of 1609, was able to build telescopes in his workshop (Figure 4-14b). In fact, Galileo was not even the first person to look at the sky through a telescope, but he was the first person to write down what he discovered and apply telescopic observations to the theoretical problem of the day: the place of Earth in the Universe.

What Galileo saw through his telescopes was so amazing that he rushed a small book into print. *Sidereus Nuncius (The Sidereal Messenger)* reported three major discoveries. First, the Moon was not perfect. It had mountains and valleys on its surface, and Galileo even used some of the mountains' shadows to calculate their height. Aristotle's philosophy held that the Moon was perfect; the "Man in the Moon" markings were believed to be reflections of continents on Earth. Galileo showed that the Moon is not only imperfect but is a world with features like Earth's.

The second discovery reported in the book was that the Milky Way was made up of myriad stars too faint to see with the unaided eye. Although intriguing, this could not match Galileo's third discovery. Galileo's telescope revealed four new "planets" circling Jupiter, objects known today as the **Galilean moons** of Jupiter (Figure 4-15).

The moons of Jupiter were strong evidence for the Copernican model. Critics of Copernicus had said Earth could not move because the Moon would be left behind, but Galileo's discovery showed that Jupiter, which everyone agreed was moving, was able to keep its satellites. This indicated that Earth, too, could move yet still hang on to the Moon. Aristotle's philosophy also included the belief that all heavenly motion was centered on Earth. Galileo's observations showed that Jupiter's moons revolve around Jupiter, which proved there could be other centers of motion besides Earth.

Sometime after *Sidereus Nuncius* was published, Galileo noticed something else that made Jupiter's moons even stronger evidence for the Copernican model. When he measured the orbital periods of the four moons, he found that the innermost moon had the shortest period and that the moons farther from Jupiter had proportionally longer periods. Jupiter's moons made up a harmonious system ruled by Jupiter, just as the planets in the Copernican universe were a harmonious system ruled by the Sun (look back to Figure 4-9b.) The similarity is not proof, but Galileo saw it as an argument that the Solar System could be Sun-centered rather than Earth-centered.



Grundy Observatory/Franklin & Marshall College/Michael A. Seeds

◀ **Figure 4-15** (a) On the night of January 7, 1610, Galileo saw three small "stars" near the bright disk of Jupiter and sketched them in his notebook. On subsequent nights (except January 9, which was cloudy), he saw that the stars were actually four moons orbiting Jupiter. (b) This photo taken through a modern telescope shows the overexposed disk of Jupiter and three of the four moons discovered by Galileo.

In the years following publication of *Sidereus Nuncius,* Galileo made two additional discoveries. When he observed the Sun, he discovered sunspots, raising the suspicion that the Sun, like the Moon, is imperfect. Further, by noting the movement of the spots, he concluded that the Sun is a sphere and that it rotates on its axis.

His most dramatic discovery came when he observed Venus. Galileo saw that it was going through phases like those of the Moon. In the Ptolemaic model, Venus moves around an epicycle centered on a line between Earth and the Sun. That means it would always be seen as a crescent (**Figure 4-16a**). But Galileo saw Venus go through a complete set of phases, which proved that it did indeed revolve all the way around the Sun (**Figure 4-16b**). There is no way the Ptolemaic model could produce the observed phases. This was the strongest evidence in support of the Copernican model that came from Galileo's telescope. Nevertheless, when controversy erupted, Galileo's critics focused more on claims about the imperfection of the Sun and Moon and the motion of the satellites of Jupiter.

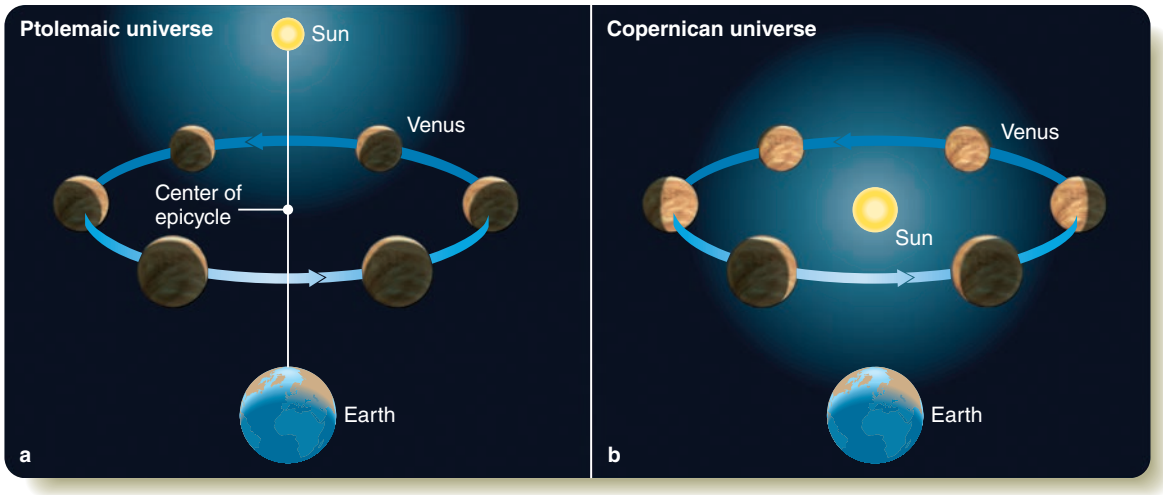
Sidereus Nuncius was popular and made Galileo famous. He became chief mathematician and philosopher to the Grand Duke of Tuscany in Florence. In 1611, Galileo visited Rome and was treated with great respect. He had long, friendly discussions with the powerful Cardinal Barberini, but he also made enemies. Galileo was outspoken, forceful, and sometimes tactless. He enjoyed argument, but most of all he enjoyed being right. In lectures, debates, and letters he offended important people who questioned his telescope discoveries.

By 1616, Galileo was the center of a storm of controversy. Some critics said he was wrong, and others said he was lying. Some refused to look through a telescope lest it mislead them, and others looked and claimed to see nothing, which is hardly surprising, given the awkwardness of those first telescopes. Pope Paul V decided to end the disruption, so when Galileo visited Rome in 1616, Cardinal Bellarmine interviewed him privately and ordered him to cease debate. There is some controversy today about the nature of Galileo's instructions, but he did not pursue astronomy for some years after the interview. Books relevant to Copernicanism were banned in all Catholic lands, although *De Revolutionibus,* recognized as an important and useful book in astronomy, was only suspended pending revision. Everyone who owned a copy of the book was required to cross out certain statements and add handwritten corrections stating that Earth's motion and the central location of the Sun were only theories and not facts. (You might recognize a similarity to current controversies about biological evolution.)

4-4b *Dialogo* and Trial

In 1621 Pope Paul V died, and his successor, Pope Gregory XV, died in 1623. The next pope was Galileo's friend Cardinal Barberini, who took the reign name of Urban VIII. Galileo rushed to Rome hoping to have the prohibition of 1616 lifted, and although the new pope did not revoke the orders, he did apparently encourage Galileo. Soon after returning home, Galileo began to write his great defense of Copernicanism, finally completing it at the end of 1629. After some delay, the book was approved by both the local censor in Florence and the head censor of the Vatican in Rome. It was printed in February 1632.

Called *Dialogo Sopra i Due Massimi Sistemi del Mondo (Dialogue Concerning the Two Chief World Systems),* it confronts the Aristotelian/Ptolemaic model with the Copernican model, using telescopic observations as evidence. Galileo wrote the book in the form of a debate among three friends: Salviati, a swift-tongued defender of Copernicus, dominates the book; Sagredo,



▲ **Figure 4-16** (a) If Venus moved in an epicycle centered on the Earth–Sun line (see Figure 3b in the Chapter 2 Concept Art: An Ancient Model of the Universe), it would always appear as a crescent. (b) Galileo observed through his telescope that Venus goes through a full set of phases, proving that it must orbit the Sun.

intelligent but largely uninformed; Simplicio, dismal defender of Ptolemy, who makes all the old arguments and sometimes doesn't seem very bright.

The publication of *Dialogo* created a storm of controversy, and it was sold out by August 1632, when the Inquisition ordered sales stopped. The book was a clear and strong defense of Copernicus, and, perhaps unintentionally, Galileo exposed papal authority to ridicule. Urban VIII was fond of arguing that, given that God is omnipotent, He could construct the Universe in any form while making it appear to humans to have a different form, and thus its true nature cannot be deduced by mere observation. Galileo placed Urban VIII's argument in the mouth of Simplicio, and Galileo's enemies showed the passage to the pope as an example of Galileo's disrespect. The pope thereupon ordered Galileo to face the Inquisition.

Galileo was interrogated by the Inquisition four times and was threatened with torture. He must have thought often of Giordano Bruno, a philosopher, poet, and Dominican monk, who was tried, condemned, and burned at the stake in Rome in 1600. One of Bruno's offenses had been advocacy of Copernicanism. However, Galileo's trial did not center on his belief in Copernicanism; *Dialogo* had been approved by two censors. Rather, the trial centered on the instructions given Galileo in 1616. From his file in the Vatican, Galileo's accusers produced a record of the meeting between Galileo and Cardinal Bellarmine that included the statement that Galileo was "not to hold, teach, or defend in any way" the principles of Copernicus. Some historians believe that this document, which was signed neither by Galileo nor by Bellarmine nor by a legal secretary, was a forgery. Others suspect it may be a draft that was never used. It is quite possible that Galileo's actual instructions were much less restrictive, but, in any case, Bellarmine was dead and could not testify at Galileo's trial.

The Inquisition condemned Galileo not for heresy but for disobeying the orders given to him in 1616. On June 22, 1633, at the age of 69, kneeling before the Inquisition, Galileo read a recantation admitting his errors. Tradition has it that as he rose he whispered, *"E pur si muove"* ("Still it moves"), referring to Earth.

Although he was sentenced to life imprisonment, he was, possibly through the intervention of the pope, confined at his villa for the next 10 years. He died there on January 8, 1642, 99 years after the death of Copernicus.

Galileo was not condemned for heresy, nor was the Inquisition concerned when he tried to defend Copernicanism. He was tried and condemned on a charge you might consider a technicality. Nevertheless, in his recantation he was forced to abandon all belief in heliocentrism. His trial has been held up as an example of the suppression of free speech and free inquiry and as a famous attempt to deny reality. Some of the world's greatest authors, including Bertolt Brecht, have written about Galileo's trial. That is why Pope John Paul II created a commission in 1979 to reexamine the case against Galileo.

To understand the trial, you must recognize that it was the result of a conflict between two ways of understanding the Universe. Since the Middle Ages, biblical scholars had taught that the only path to true understanding was through religious faith. St. Augustine (354–430) wrote *"Credo ut intelligam,"* which can be translated as "Believe in order to understand." Galileo and other scientists of the Renaissance, however, used their own observations as evidence to try to understand nature. When their observations contradicted Scripture, they assumed that their observations represented reality. Galileo paraphrased Cardinal Baronius in saying, "The Bible tells us how to go to heaven, not how the heavens go." The trial of Galileo was not really about the place of Earth in the Universe. It was not about Copernicanism. It wasn't even about the instructions Galileo received in 1616. It was, in a larger sense, about the birth of modern science as a rational way to understand the physical Universe.

The commission appointed by Pope John Paul II in 1979, reporting its conclusions in October 1992, said of Galileo's inquisitors, "This subjective error of judgment, so clear to us today, led them to a disciplinary measure from which Galileo 'had much to suffer.'" Galileo was not found innocent in 1992 so much as the clerical judges were deemed, 400 years afterward, of having failed to fully grapple with the real questions involved.

Practicing **SCIENCE**

How were Galileo's observations of the moons of Jupiter evidence against the Ptolemaic model? In this case, the distinction scientists usually make between model and hypothesis is blurred.

The Ptolemaic model was considered to be more than just a method for calculating planetary motions. That model was understood to represent the Aristotelian universe, believed to be true by virtually all astronomers and philosophers. The Ptolemaic model implied predictions about what the first astronomer with a telescope should see. Making predictions that can be used to check and compare hypotheses, as well as making observations to check predictions of hypotheses, are both central to practicing science.

Galileo presented his scientific reasoning in the form of evidence and conclusions; the moons of Jupiter were some of his key evidence. Moons circling Jupiter did not fit the implicit Aristotelian belief that all motion is centered on Earth. Obviously, there can be other centers of motion.

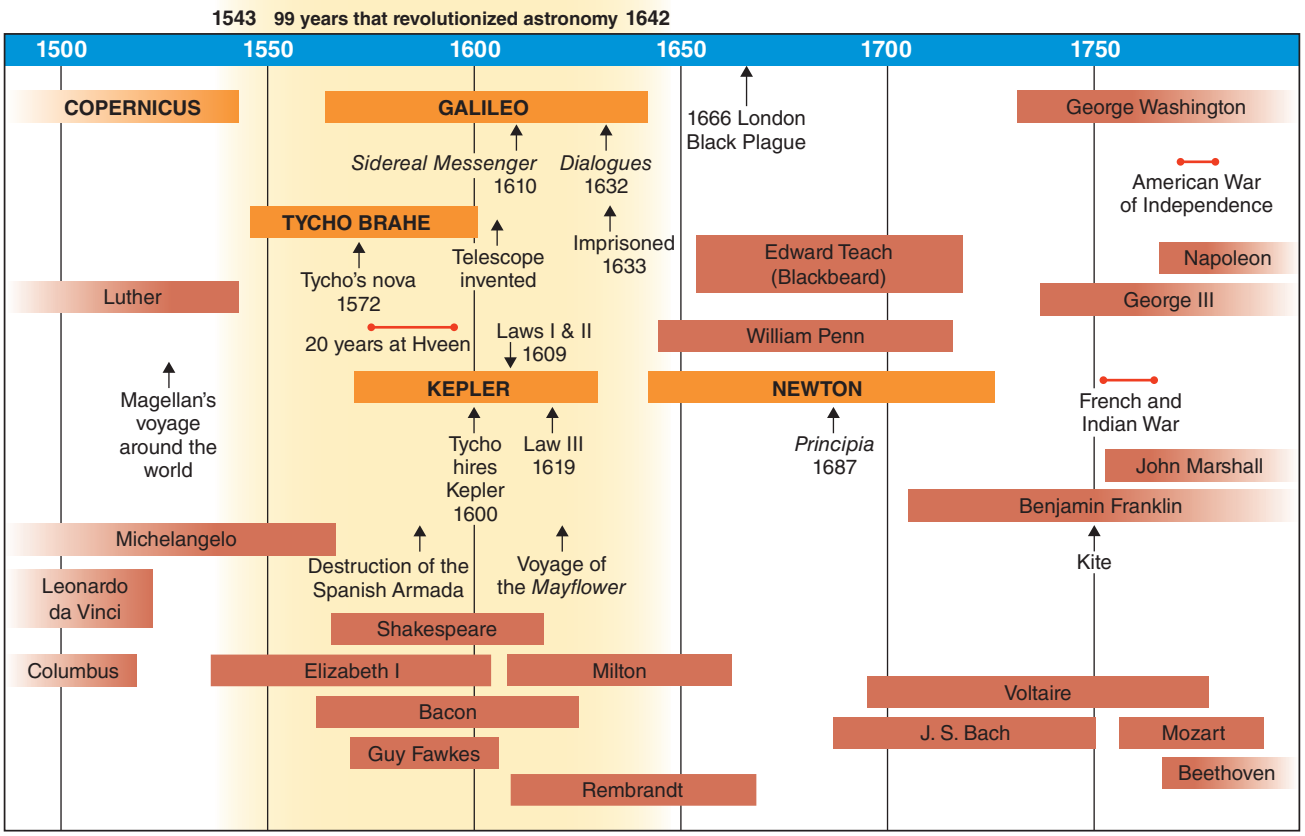
The phases of Venus provide even better evidence against the Ptolemaic model and for the Copernican model. Both models imply specific, and very different, predictions about how the phases of Venus should appear. In the Ptolemaic model, Venus will always be seen as a crescent from Earth. In the Copernican model, Venus will go through a Moon-like sequence of phases, from crescent to full and back to crescent, and the planet should have the smallest angular size when it is at full phase. What Galileo saw was exactly what the Copernican model would predict.

4-5 Ninety-Nine Years That Revolutionized Astronomy

The transition from ancient to modern astronomy began during the 99 years between the publication of *De Revolutionibus* (1543) and Galileo's death (1642). That transition began with the replacement of the Ptolemaic model of the Universe by the Copernican model, with a closely related controversy over the place of Earth in the Universe. That same period also saw an evolution in the methods of science in general, illustrated by the solution of the puzzle of planetary motion. That puzzle was not solved by philosophical arguments about the perfection of the heavens or by debate over the meanings of scripture. It was solved by precise observations and careful computations, techniques that are the foundations of modern science.

The discoveries made by Kepler and Galileo found acceptance in the 1600s because the world was in transition. Astronomy was not the only thing changing during this period. The Renaissance is commonly taken to be the period between about the years 1300 and 1600; these 99 years of astronomical history thus lie at the culmination of a reawakening of learning in all fields (**Figure 4-17**). Ships were sailing to new lands and encountering new cultures. The world was open to new ideas and new observations. Martin Luther started a reformation of the Christian religion, and other philosophers and scholars rethought their respective areas of human knowledge. Had Copernicus not published his hypothesis, someone else would have suggested that the Solar System is heliocentric. History was ready to shed Aristotle's universe and the Ptolemaic model.

Beginning with Copernicus, scientists such as Tycho, Kepler, and Galileo depended more and more on evidence, observation, and measurement rather than on first principles. That change was connected with the Renaissance via advances in metalworking and lens making. At the time of Copernicus, no astronomer had looked through a telescope because one could not be made. By 1642, not only telescopes but also other sensitive measuring instruments had transformed science into something new and precise. As you can imagine, scientists were excited by these discoveries, and they founded scientific societies that increased the exchange of observations and hypotheses, and stimulated more and better work. The most important advance, however, was the application of mathematics to scientific questions. Kepler's work demonstrated the power of mathematical analysis, and as the quality of these numerical techniques improved, the progress of science accelerated. The story of the birth of modern astronomy is actually the story of the birth of modern science as well.



▲ **Figure 4-17** The 99 years between the death of Copernicus in 1543 and the death of Galileo in 1642 marked the transition from the ancient astronomy of Ptolemy and Aristotle to the revolutionary theory of Copernicus, and, simultaneously, the invention of science as a way of understanding nature.

WHAT ARE WE?

Thinkers

The scientific revolution began when Copernicus made humanity part of the Universe. Before Copernicus, people thought of Earth as a special place different from any of the objects in the sky, but in trying to explain the motions in the sky, Copernicus made Earth one of the planets. Galileo and those who brought him to trial understood the significance of making Earth a planet. It made Earth and humanity part of nature, part of the Universe.

Kepler showed that the planets move according to simple rules. We are not in a special place ruled by mysterious planetary forces. Earth, the Sun, and all of humanity are part of a Universe in which motions can be described by a few fundamental laws. If simple laws describe the motions of the planets, then the Universe is not ruled by mysterious influences as in astrology or the whims of the gods atop Mount Olympus. And if the Universe can be described by simple rules, then it is open to scientific study.

Before Copernicus, people felt they were special because they thought they were at the center of the Universe. Copernicus, Kepler, and Galileo showed that we are not at the center but nevertheless are part of an elegant and complex Universe. Astronomy tells us that we are special because we can study the Universe and eventually understand what we are. It also tells us that we are not only part of nature; we are the part of nature that can think about nature.

Study and Review

SUMMARY

4-1 Roots of Astronomy

- **Archaeoastronomy** is the study of the astronomical knowledge of ancient peoples. Many cultures around the world observed the sky and marked important alignments, though some left no written record of their astronomical beliefs.
- Classical philosophers accepted as a **first principle** that Earth was the unmoving center of the Universe. Another first principle was that the heavens were perfect, so philosophers such as Plato argued that, because the sphere was the most perfect geometrical form, the heavens must be made up of spheres in uniform rotation. This led to the belief in **uniform circular motion**.
- Many astronomers argued that Earth could not be moving because they could see no **parallax** in the positions of the stars.
- Aristotle's estimate for the size of Earth was only about 40 percent of its true size. Eratosthenes used sunlight shining to the bottom of a well at the southern Egyptian city of Syene to measure the diameter of Earth and produced an accurate estimate.
- The **geocentric universe** became part of the teachings of the great philosopher Aristotle, who argued that the Sun, Moon, and stars were carried around Earth on rotating crystalline spheres.
- Hipparchus, who lived about two centuries after Aristotle, devised a model in which the Sun, Moon, and planets revolved in circles called **eccentrics** with Earth near, but not precisely at, their centers.
- **Retrograde motion**, the occasional westward (backward) motion of the planets, was difficult for astronomers to explain.
- About the year 140, Aristotle's model was given mathematical form in Claudius Ptolemy's book *Almagest*. Ptolemy preserved the principles of geocentrism and uniform circular motion, but he added **epicycles**, **deferents**, and **equants**. Ptolemy's epicycles could approximate retrograde motion, but the Ptolemaic model was not accurate, and it had to be revised a number of times as centuries passed.

4-2 The Copernican Revolution

- Copernicus devised a **heliocentric universe**. He preserved the principle of uniform circular motion, but he argued that Earth rotates on its axis and revolves around the Sun once a year. His theory was controversial because it contradicted Church teaching about Earth's place in the Universe. Copernicus published his theory in his book *De Revolutionibus* in 1543, the same year he died.
- A **hypothesis** is a specific statement about nature that needs further testing, but a **theory** is usually a general description of some aspect of nature that is well understood, has been thoroughly tested, and is widely accepted. A **natural law** is a fundamental and universal principle in which scientists have great confidence.
- Because Copernicus kept uniform circular motion, his model did not offer improved predictions of planetary motions, but it did offer a simple explanation of retrograde motion without using big epicycles.
- One reason the Copernican model won converts was that it was more elegant than the Ptolemaic model. Venus and Mercury were treated the same as all the other planets, and the speed of each planet was related to its distance from the Sun.
- The shift from the geocentric **paradigm** to the heliocentric paradigm is an example of a scientific revolution.

4-3 Tycho, Kepler, and Planetary Motion

- Although Tycho Brahe developed his own model in which the Sun and Moon circled Earth and the planets circled the Sun, his great contribution was to compile detailed, precise observations of the positions of the Sun, Moon, and planets over a period of 20 years, observations that were later used by Johannes Kepler.
- Kepler inherited Tycho's books of observations in 1601 and used them to discover three laws of planetary motion. He found that the planets follow orbits that are **ellipses** with the Sun at one focus, that they move faster when near the Sun, and that a planet's orbital period squared is proportional to the **semimajor axis, *a***, of its orbit cubed.

$$P_{yr}^2 = a_{AU}^3 \qquad \text{(Eq. 4-1)}$$

These laws are all **empirical**, describing the phenomena of planetary motion without supplying explanations.

- The **eccentricity, *e***, of an orbit is a measure of its departure from a perfect circle. A circle is an ellipse with an eccentricity of zero.
- Kepler's final book, the *Rudolphine Tables* (1627), combined heliocentrism with elliptical orbits and predicted the positions of the planets 10 to 100 times more accurately than any previous effort.

4-4 Galileo's Conclusive Evidence

- Galileo used the newly invented telescope to observe the heavens, and he recognized the significance of what he saw there. His discoveries of the phases of Venus, the satellites of Jupiter now known as the **Galilean moons**, the mountains of Earth's Moon, and other phenomena helped undermine the Ptolemaic universe.
- Galileo based his analysis on observational evidence. In 1633, he was condemned by the Inquisition for disobeying instructions not to hold, teach, or defend Copernicanism.
- Historians of science view Galileo's trial as a conflict between two ways of knowing about nature, reasoning from first principles versus depending on evidence.

4-5 Ninety-Nine Years That Revolutionized Astronomy

- The 99 years from the death of Copernicus to the death of Galileo marked the birth of modern science. From that time on, science depended on evidence to test theories and relied on the mathematical analytic methods first demonstrated by Kepler.

REVIEW QUESTIONS

1. What evidence suggests that early human cultures observed astronomical phenomena?
2. Why did early human cultures observe astronomical phenomena? Was it for scientific research?
3. Early cultures believed that ultimate causes of things are mysteries beyond human understanding. Is this a first principle? Why or why not?
4. Name one example each of a famous politician, mathematician, philosopher, observer, and theoretician in this chapter.
5. Why did Plato propose that all heavenly motion was uniform and circular?
6. On what did Plato base his knowledge? Was it opinions, policies, marketing, public relations, myths, evidence, hypotheses, beliefs, laws, principles, theories? On what do modern astronomers base their knowledge?
7. Which two-dimensional (2D) and three-dimensional (3D) shapes did Plato and Aristotle consider perfect? Give an example of a 2D and 3D nonperfect geometrical shape.
8. Are the spheres of Eudoxus a scientific model? If so, is it entirely true?
9. In Ptolemy's model, how do the epicycles of Mercury and Venus differ from those of Mars, Jupiter, and Saturn?
10. Describe in detail the motions of the planets according to Ptolemy.
11. In Ptolemy's model, which of the following—epicycle, equant, or deferent—travels in uniform circular motion as viewed from a particular point? Name and describe that point. Are these uniform circular motions at the same speeds and in the same directions?
12. Why did Copernicus have to keep small epicycles in his model? Which planet has the longest duration of retrograde motion as viewed from Earth? The shortest?
13. Was the belief held by ancient astronomers that the Moon and Sun are unblemished a paradigm, a scientific revolution, or neither?
14. When Tycho observed the new star of 1572, he could detect no parallax. Why did that undermine belief in the Ptolemaic system? In the perfect heavens idea of Aristotle?
15. Assume the night is clear and the Moon's phase is full so you can see it all night long. Why would you expect the Moon to show parallax during the night? How would you measure it?
16. Does Tycho's model of the Universe explain the phases of Venus that Galileo observed? Why or why not?
17. Name an empirical law. Why is it considered empirical?
18. How does Kepler's first law of planetary motion overthrow one of the basic beliefs of classical astronomy? How about Kepler's second law?
19. When Mercury is at aphelion (farthest from the Sun) in Figure 4-13a, compare *a,* the semimajor axis, to *r,* the distance from the Sun to Mercury. Is *a* greater than *r,* or less than *r,* or equal to *r?*
20. In Figure 4-13b, did the planet travel faster from points *A* to *B,* travel faster from points *A′* to *B′,* travel the same speed between *A* to *B* and *A′* to *B′,* or can you not determine speeds of a planet's orbit based on the information given? How do you know?
21. What is *P* for Earth? What is *a* for Earth? Do these values support or disprove Kepler's third law?
22. Based Figure 4-13c, do planets with larger *a* take longer, shorter, or the same time to orbit the Sun?
23. How did the *Alfonsine Tables,* the *Prutenic Tables,* and the *Rudolphine Tables* differ?
24. Explain how each of Galileo's telescopic discoveries contradicted the Ptolemaic theory.
25. How did discovery of the Galilean moons disprove Plato's and Aristotle's perfect heavens first principle(s)?
26. **How Do We Know?** What is a paradigm, and how it is related to a scientific revolution? Give an example of a paradigm and a scientific revolution.
27. **How Do We Know?** Describe the differences between a hypothesis, a theory, and a law. Give an example of each.

ACTIVE INQUIRY QUESTIONS

1. Science historian Thomas Kuhn has said that *De Revolutionibus* was a revolution-making book but not a revolutionary book. How was that book in some ways still classical and conservative?
2. Many historians suspect that Galileo offended Pope Urban VIII by putting the pope's favorite argument into the mouth of Simplicio. How is the pope's argument a challenge to the principle of cause and effect?
3. How does the modern controversy over creationism and evolution reflect two ways of knowing about the physical world?

PROBLEMS

1. Draw and label a diagram of the western horizon from northwest to southwest and label the setting points of the Sun at the solstices and equinoxes for a person in the Northern Hemisphere. (*Hint:* See Figure 3 in the Chapter 2 Concept Art: **The Sky Around You** and Figure 4-1.)
2. If you lived on Mars, which planets would exhibit retrograde motion like that observed for Mars from Earth? Which would never be visible as crescent phases?
3. How long does it take for one retrograde cycle of Mars as viewed from Earth, and in which direction is the retrograde motion? What fraction of Mars's orbit around the Sun is the duration of retrograde motion as viewed from Earth?

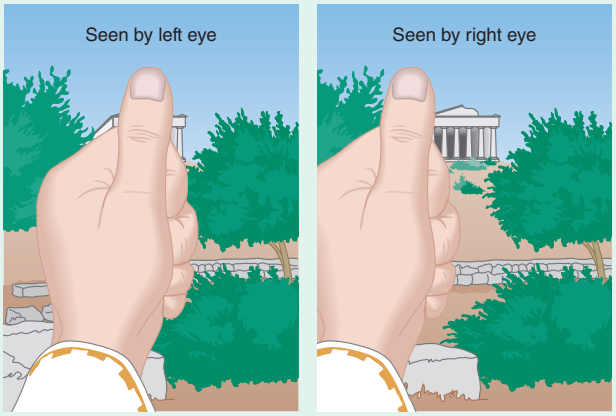
4. If a planet has an average distance from the Sun of 2.0 AU, what is its orbital period?
5. If a space probe is sent into an orbit around the Sun that brings it as close as 0.4 AU and as far away as 5.4 AU, is the orbit a circle or an ellipse? What will be its orbital period?
6. Uranus orbits the Sun with a period of 84.0 years. What is its average distance from the Sun?
7. An object takes 29.5 years to orbit the Sun. What is its average distance from the Sun?
8. One planet is three times farther from the Sun than another. Will the farther planet take more, less, or the same amount of time to orbit the Sun? Will the closer planet orbit slower, faster, or the same speed? How much longer will the farther planet take to orbit than the closer planet? If the closer planet is located at 10 AU, how far is the farther planet, and what are the two planet's names?
9. Galileo's telescope showed him that Venus has a large angular diameter (61 arc seconds) when it is a crescent and a small angular diameter (10 arc seconds) when it is nearly full. Use the small-angle formula to find the ratio of its maximum to minimum distance from Earth. Is this ratio compatible with the Ptolemaic universe shown in Figure 3b of the Chapter 4 Concept Art: **An Ancient Model of the Universe?**
10. Which is the phase of Venus when it is closest? Which when farthest? How do you know?
11. Galileo's telescopes were not of high quality by modern standards. He was able to see the moons of Jupiter, but he never reported seeing features on Mars. Use the small-angle formula to find the angular diameter of Mars when it is closest to Earth. How does that compare with the maximum angular diameter of Jupiter?

SENSE OF PROPORTION

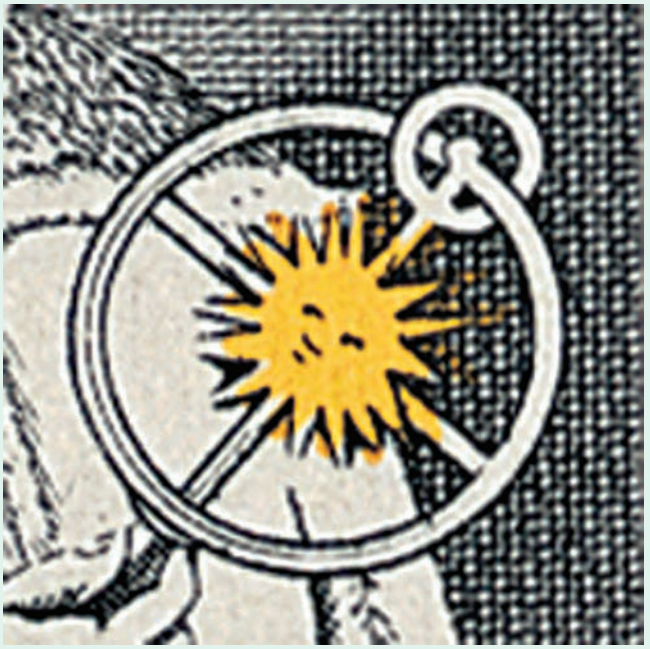
1. Arrange the following in order of increasing distance from Earth according to the geocentric model of the Universe: Jupiter, Mercury, the Sun, Venus, Mars, Saturn, the stars
2. Arrange the following in order of increasing distance from the Sun according to the heliocentric model of the Universe: Jupiter, Mercury, Earth, Venus, Mars, Saturn, the stars

LEARNING TO LOOK

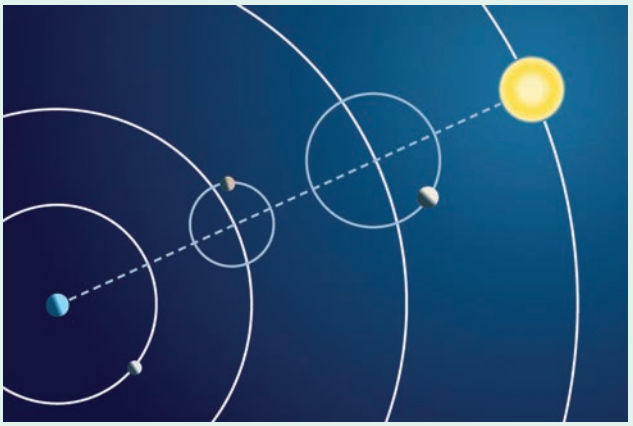
1. With an outstretched arm, hover your thumb about ½ inch above the page and over the thumb shown in the "Seen by right eye" image in part 1a of **An Ancient Model of the Universe**. Close one eye and rapidly blink closed one eye then the other and repeat several more times. Why does the building in the picture not disappear as shown in the "Seen by left eye" picture? Is this an example of seeing parallax? What do you need to do to see a very large parallax?



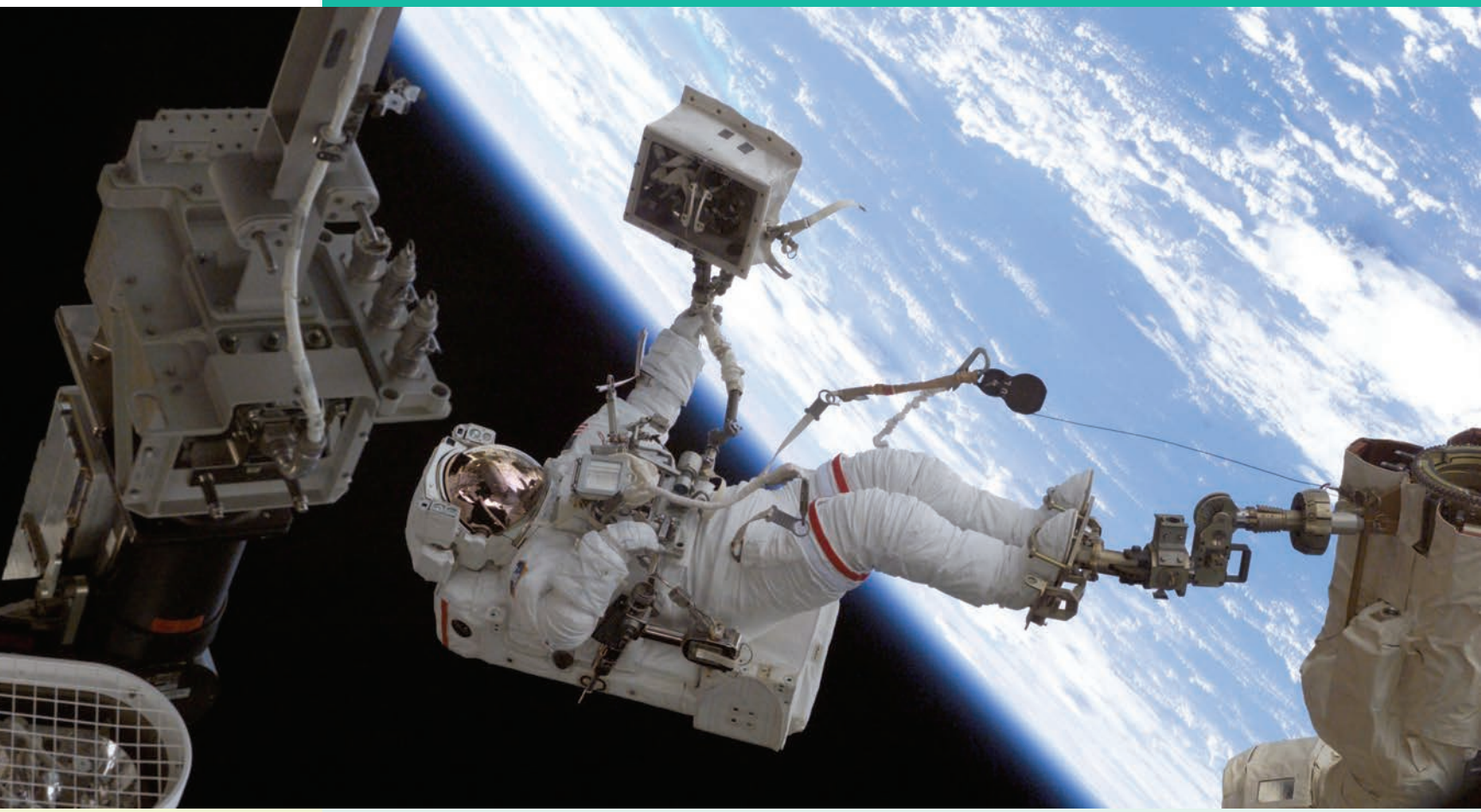
2. Study Figures 4-11 and 4-16 and describe the phases that Venus would have displayed to Galileo's telescope if the Tychonic universe had been correct.


© 2016 Cengage

3. What three astronomical objects are represented here? What are the two rings?
4. Use the figure below to explain how the Ptolemaic model treated some planets differently from the rest. How did the Copernican model treat all of the planets the same?



5 Gravity



NASA

▲ Space stations and astronauts, as well as planets, moons, stars, and galaxies, follow paths called *orbits* that are described by three simple laws of motion and a theory of gravity first proposed by Isaac Newton.

If Renaissance astronomers had understood gravity, they would have had much less trouble understanding the motions of the planets, but real insight about gravity didn't come until three decades after Galileo's trial. Isaac Newton started from Galileo's work and devised a way to explain motion and gravity that allowed astronomers to understand orbits and tides with great precision. Later, in the early 20th century, Albert Einstein found an even better way to explain motion and gravity that included and broadened Newton's previous explanations.

This chapter is about gravity, the master of the Universe. Here you will find answers to five important questions:

- What were Galileo's insights about motion and gravity?
- What did Newton add to the understanding of motion and gravity?
- How does gravity explain orbital motion?
- How does gravity explain tides?
- How did Einstein change conceptions of motion and gravity?

Gravity rules. From the Moon orbiting Earth, to matter falling into black holes, to the formation of galaxy clusters, as you study the Universe you will see gravity in action everywhere.

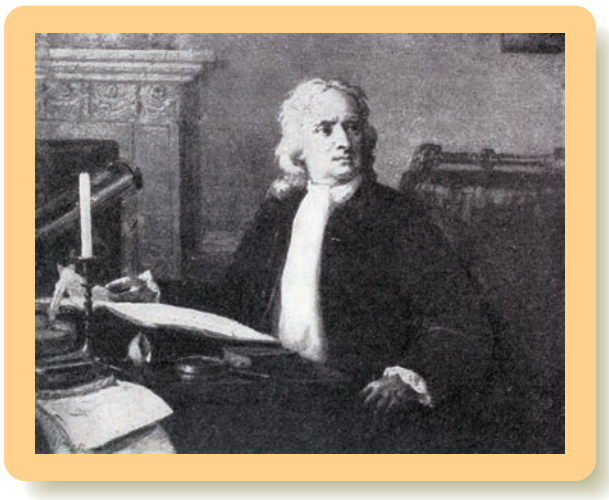
> **Nature and Nature's laws lay hid in night: God said, "Let Newton be!" and all was light.**
>
> ALEXANDER POPE

DOESN'T IT SEEM strange that Isaac Newton (1642–1727; Figure 5-1) is said to have "discovered" gravity in the late 17th century—as if people didn't have gravity before that, as if they floated around holding onto tree branches? Of course, everyone experienced gravity while taking it for granted. Newton's insight was to see that the force of gravity that makes apples fall to Earth also keeps moons and planets in their orbits. That realization changed the way people thought about nature.

5-1 Galileo's and Newton's Two New Sciences

Newton was born in Woolsthorpe, England, on December 25, 1642, and on January 4, 1643. This was not a biological anomaly but a quirk of the calendar. Most of Europe, following the lead of the Catholic countries, had adopted the new Gregorian calendar, but Protestant England continued to use the older Julian calendar. So December 25 in England was January 4 in Europe. If you use the English date, then Newton was born in the same year that Galileo died.

Newton became one of the greatest scientists in history, but he realized the debt he owed to those who had studied nature before him. Newton said, "If I have seen farther it is by standing on the shoulders of giants." One of those giants was Galileo. In the previous chapter you learned that Galileo was the great defender of Copernicanism who made the first recorded use of an astronomical telescope, but he was also the first scientist who carefully studied the motions of falling bodies. Galileo provided the key information that helped Newton understand gravity.



iryna1/Shutterstock.com

▲ **Figure 5-1** Isaac Newton proposed three simple laws of motion and a theory of gravity that together determine the motions of spacecraft, moons, planets, stars, and galaxies.

5-1a Galileo's Observations of Motion

Galileo was performing experiments to study forces and motion years before he built his first telescope in 1609 (look back to Section 4-4). After the Inquisition condemned and imprisoned him in 1633, he continued his study of motion. He seems to have realized that he needed to understand motion before he could truly comprehend the Copernican system.

In addition to writing about a geocentric universe that you studied in the previous chapter, Aristotle also wrote about the nature of motion, and those ideas still held sway in Galileo's time. Aristotle said that the world is made up of four elements: earth, water, air, and fire, with each located in its proper place. (Ironically, it is accurate to think of living creatures especially as being really composed of a combination of earth = soil, water, air, and fire = energy.) The proper place for earth (meaning soil and rock) is the center of the Universe, and the proper place of the water element is just above the earth element. Air and then fire form higher layers, and above them is the realm of the planets and stars, made of a celestial substance unknown on Earth. (The Earthly elements are shown as layers in the center of the diagram in Figure 1 of the Chapter 4 Concept Art: An Ancient Model of the Universe.)

Aristotle wrote, and so everyone believed for almost 2000 years afterward, that objects have a natural tendency to move toward their proper places in the cosmos. Things made up mostly of air or fire—smoke, for instance—tend to move upward. Things composed mostly of earth and water—wood, rock, flesh, bone, and so on—tend to move downward. Therefore, according to Aristotle, objects that fall downward do so because they are moving toward their proper place. That is one reason Aristotle's universe had to be geocentric. His explanation of gravity—why things fall down—works only if the center of Earth is also the center of the Universe.

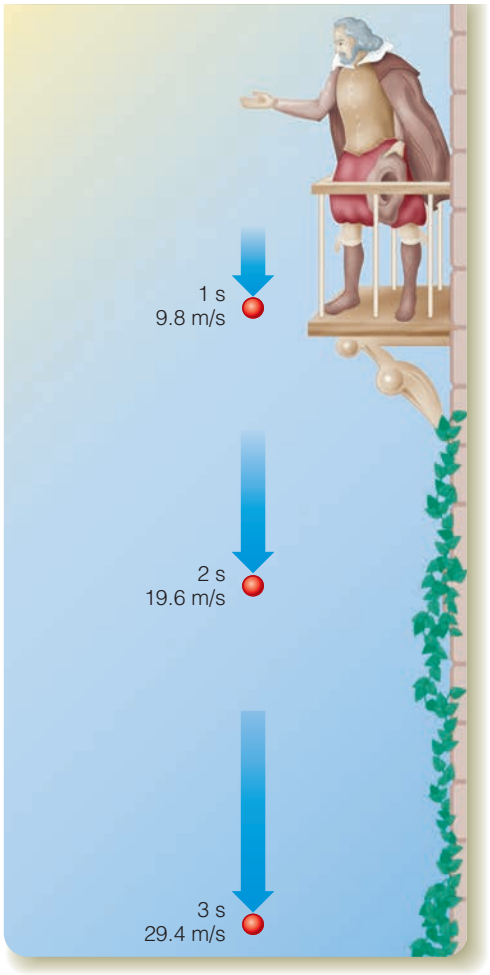
Aristotle called these motions **natural motions**; this was to distinguish them from **violent motions** that are produced when, for instance, you push on an object and force it to move other than toward its proper place. According to Aristotle, such motions stop as soon as the force is removed. To explain how an arrow could continue to move upward even after it had left the bowstring, he said that currents in the air around the arrow carried it forward even though the bowstring was no longer pushing it.

In Galileo's time, as well as during the two preceding millennia, scholars had tended to resolve problems by referring to authority. To analyze the flight of a cannonball, for instance, they would turn to the writings of Aristotle and other classical philosophers and try to deduce what those philosophers would have said on the subject. This generated a great deal of discussion

but little real progress. Galileo broke with this tradition when he conducted his own experiments and, furthermore, believed the results were more informative than ancient authority.

Galileo began by studying the motions of falling bodies, but he quickly discovered that the velocities were so great and the times so short that he could not measure them accurately. To solve that problem, Galileo began using polished metal balls rolling down gently sloping inclines so the velocities were lower and the times longer. Using an ingenious water clock he invented, Galileo was able to measure the amount of time the balls took to roll given distances down the incline and, most important, correctly recognized that those times are proportional to the times he would have measured for freely falling bodies.

Galileo found that falling bodies do not fall at constant rates, as Aristotle had said, but are accelerated. That is, they move faster with each passing second. Near Earth's surface, a falling object will have a velocity of 9.8 m/s (32 ft/s) at the end of 1 second, 19.6 m/s after 2 seconds, 29.4 m/s after 3 seconds, and so on. Each passing second adds 9.8 m/s (32 ft/s) to the object's velocity (**Figure 5-2**). In modern terms, this steady increase in the velocity of a falling body by 9.8 m/s each second (usually written 9.8 m/s^2, read as "9.8 meters per second squared") is called the **acceleration of gravity** at Earth's surface.



▲ **Figure 5-2** Galileo found that a falling object is accelerated downward. Each second its velocity increases by 9.8 m/s (32 ft/s).

NASA

▲ **Figure 5-3** Air resistance would confuse the result of an experiment like the one attributed to Galileo, dropping objects of equal size but different masses from a high tower to compare their rates of fall. In a low-resolution television broadcast from the airless surface of the Moon in 1971, Apollo 15 commander David Scott dropped a hammer and a feather at the same moment. With no air resistance, they fell with the same acceleration and hit the ground together. The same demonstration can be done more prosaically in a laboratory or classroom on Earth using a container with the air pumped out.

Galileo also discovered that the acceleration does not depend on the weight of the object. This, too, is contrary to the teachings of Aristotle, who believed that heavy objects, containing more earth and water, fell with higher velocity. Galileo found that the acceleration of a falling body is the same whether it is heavy or light. According to some accounts, he demonstrated this by dropping balls of iron and wood from the top of the Leaning Tower of Pisa to show that they would fall together and hit the ground at the same time. In fact, he probably did not perform this experiment. It would not have been conclusive anyway because of the effect of air resistance. More than 300 years later, Apollo 15 astronaut David Scott, standing on the airless Moon, demonstrated the truth of Galileo's discovery by simultaneously dropping a feather and a steel geologist's hammer. They fell at the same rate and hit the lunar surface at the same time (**Figure 5-3**).

Having described natural motion, Galileo turned his attention to what Aristotle called "violent" motion—that is, motion directed other than toward an object's proper place in the cosmos. Aristotle said that such motion must be sustained by a cause. Today we would say "sustained by a force." Galileo pointed out that an object rolling down an incline is accelerated and that an object rolling up the same incline is decelerated. If

the incline were perfectly horizontal and frictionless, he reasoned, there could be no acceleration or deceleration to change the object's velocity, and in the absence of friction, the object would continue to move forever. In Galileo's own words, "Any velocity once imparted to a moving body will be rigidly maintained as long as the external causes of acceleration or retardation are removed." In other words, motion does not need to be sustained by a force, said Galileo, disagreeing with Aristotle. Once begun, motion continues until something changes it. This property of matter is called **inertia**. In fact, Galileo's description of what is essentially a law of inertia is a good summary of the principle that became known as Newton's first law of motion.

Galileo published his work on motion in 1638, two years after he had become entirely blind and only four years before his death. The book was named *Discourses and Mathematical Demonstrations Concerning Two New Sciences, Relating to Mechanics and to Local Motion* (in Italian, *Discorsi e Dimostrazioni Matematiche, Intorno à Due Nuove Scienze, Attenenti alla Mecanica & i Movimenti Locali*). It is known today as *Two New Sciences.*

The book is a brilliant achievement for a number of reasons. To understand motion, Galileo had to abandon ancient authority, devise his own experiments, and draw his own conclusions. In a sense Galileo's work was the first example of modern experimental science. Also notice that Galileo was able to make valid general conclusions about how nature works based on his limited experiments. Though his apparatus was finite and his results affected by friction, he was able to imagine an infinite, frictionless plane on which a body moves at constant velocity. In his workshop, the law of inertia was obscure, but in his imagination it was clear and precise.

5-1b Newton's Laws of Motion

From the work of Galileo, Kepler, and other early scientists, Newton was able to deduce three laws of motion (**Table 5-1**) that describe any moving object, from an automobile driving along a highway to galaxies colliding with each other. Those laws of motion led Newton further, to an understanding of gravity.

Newton's first law of motion is really a restatement of Galileo's law of inertia: An object continues at rest or in uniform motion in a straight line unless acted on by some force. For example, an astronaut drifting in space will travel at a constant speed in a straight line forever if no forces act on her (**Figure 5-4a**).

Newton's first law also explains why a projectile continues to move after all forces have been removed—for instance, how an arrow continues to move after leaving the bowstring. The object continues to move because it has **momentum**.

An object's momentum is a measure of its amount of motion, equal to its velocity multiplied by its **mass**. (Mass is a measure of the amount of matter in an object, expressed in kilograms in the metric system.) A paper clip tossed across a room has low velocity and therefore little momentum, and you could easily catch it in your hand. But the same paper clip fired at the speed of a rifle bullet would have tremendous momentum, and you would not dare try to catch it. Momentum also depends on the mass of an object. Now imagine that, instead of tossing a paper clip, someone tossed you a bowling ball. A bowling ball contains much more mass than a paper clip and therefore has much greater momentum, even though it is moving with the same velocity. If you tried to catch a bowling ball moving as fast as the tossed paper clip, it might knock you down.

Note that mass is not the same as weight. Your weight is the force that Earth's gravity exerts on the mass of your body. Because gravity pulls you downward, you press against the bathroom scale, and you can measure your weight. Floating in space, you would have no weight at all; a bathroom scale would be useless. Nevertheless, your body would still contain the same amount of matter, so you would still have the same mass you do on Earth. The same would be true for the bowling ball; imagine throwing one in space, where it would be weightless. It would still be massive, and substantial effort would be required to start it moving.

Newton's second law of motion is about forces. Where Galileo spoke only of accelerations, Newton saw that acceleration is the result of force acting on a mass (**Figure 5-4b**). Newton's second law is commonly written as:

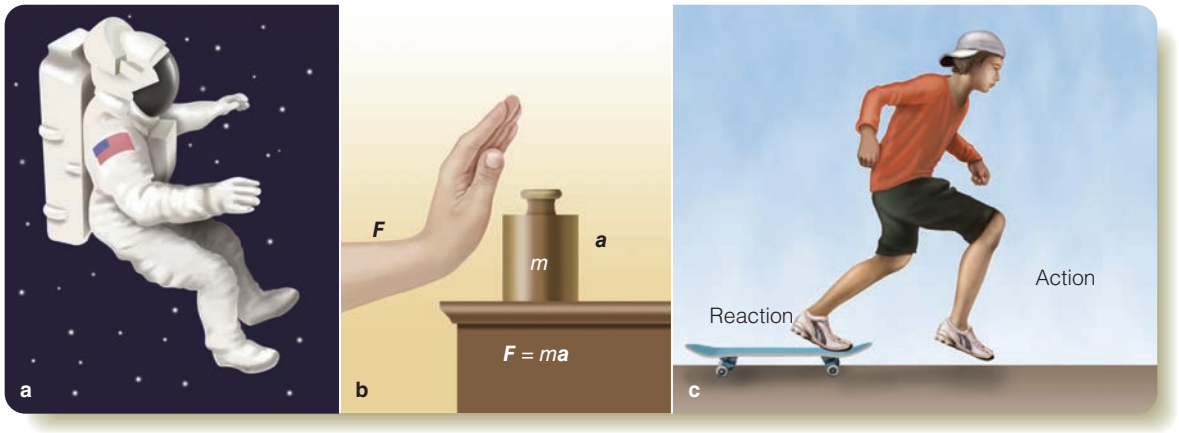
$$F = ma$$

where F is force, m is mass, and a is acceleration. As always, you need to define terms carefully when you look at an equation. **Acceleration** is a change in velocity, and **velocity** is speed with a specific direction. Most people use the words *speed* and *velocity* interchangeably, but they mean two different things. Speed is a rate of motion and does not have any direction implied, but velocity does. If you drive a car in a circle at 55 mph, your speed is constant, but your velocity is changing because your direction of motion is changing. An object experiences acceleration if its speed changes or *if its direction of motion changes.* Every automobile has three accelerators—the gas pedal, the brake pedal, and the steering wheel. All three change the car's velocity, in the technical sense of the word.

In a way, the second law is just common sense; you experience its consequences every day. The acceleration of a body is proportional to the force applied to it. If you push gently against a grocery cart, you expect a small acceleration. The second law of motion also says that the acceleration depends on the mass of the body. If your grocery cart is filled with bricks and you push it gently, you expect little result. If it is full of inflated balloons, however, it begins accelerating easily in response to a gentle

Table 5-1 Newton's Three Laws of Motion

I.	A body continues at rest or in uniform motion in a straight line unless acted on by an external force.
II.	The acceleration of a body is inversely proportional to its mass, directly proportional to the force, and in the same direction as the force.
III.	To every action, there is an equal and opposite reaction.



▲ **Figure 5-4** Illustrations of Newton's three laws of motion. (a) First law: All objects, such as the astronaut pictured, continue at rest or move at a constant speed in a straight line unless acted on by some force. (b) Second law: The force needed to accelerate an object is proportional to the mass of the object and the acceleration. (c) Third law: The force exerted on the skateboard by the skateboarder is equal and opposite to the force exerted on the skateboarder by the skateboard.

push. Finally, the second law says that the resulting acceleration is in the direction of the force. This is also what you would expect. If you push on a cart that is not moving, you expect it to begin moving in the direction you push.

The second law of motion is important because it establishes a precise relationship between cause and effect (**How Do We Know? 5-1**). Objects do not just move; they accelerate as a result of the action of a force. Moving objects do not just stop; they decelerate as a result of a force. And moving objects don't just change direction for no reason. A change in direction is a change in velocity and requires the presence of a force. Aristotle said that when objects move, it is because they have a tendency to move to their proper places. Newton said that objects move because of a specific cause, a force. And Newton's second law is in the form of an equation, so, for example, you can calculate an object's precise numerical amount of acceleration if you know the numerical amounts of its mass and the force acting on it.

Newton's third law of motion specifies that for every action there is an equal and opposite reaction. In other words, forces must occur in pairs directed in opposite directions. For example, if you stand on a skateboard and jump forward, the skateboard will shoot away backward. As you jump, your feet must exert a force against the skateboard, which accelerates it toward the rear. But, as Newton realized, forces must occur in pairs, so the skateboard must exert an equal but opposite force on your feet, and that is what accelerates your body forward (**Figure 5-4c**).

5-1c Mutual Gravitation

Once Newton understood the three laws of motion, he was able to consider the force that causes objects to fall. The first and second laws tell you that falling bodies accelerating downward means there must be some force pulling downward on them. In Aristotle's view, the Moon and other celestial bodies move perpetually in circles because that is the nature of whatever heavenly (un-Earthly) substance they are composed. Newton assumed instead that the Moon and other celestial bodies have the same nature and follow the same rules as objects on Earth, which means that some force must act on the Moon to keep it in orbit. The Moon follows a curved path around Earth, and motion along a curved path is accelerated motion. The second law of motion says that acceleration requires a force, so a force must be making the Moon follow that curved path.

Newton wondered if the force that holds the Moon in its orbit could be the same force that causes apples to fall—gravity. He was aware that gravity extends at least as high as the tops of mountains, but he did not know if it could extend all the way to the Moon. He believed that it could, but he thought it would be weaker at greater distances, and he assumed that its strength would decrease as the square of the distance increased.

This relationship, the **inverse square law**, was familiar to Newton from his work on optics, where it applied to the intensity of light. A screen set up 1 meter from a candle flame receives a certain amount of light on each square meter. However, if that screen is moved to a distance of 2 meters, the light that originally illuminated 1 square meter must now cover 4 square meters (**Figure 5-5**). Consequently, the intensity of the light is inversely proportional to the square of the distance to the screen. Newton guessed that the intensity of gravity would follow the same law.

Newton made a second assumption that enabled him to predict the strength of Earth's gravity at the distance of the Moon. He assumed not only that the strength of gravity follows the inverse square law but also that the important distance to consider is the distance from Earth's center, not the distance from Earth's surface. (By the way, Newton invented calculus to verify that assumption.) Because the Moon is about 60 Earth radii away, Earth's gravity at the distance of the

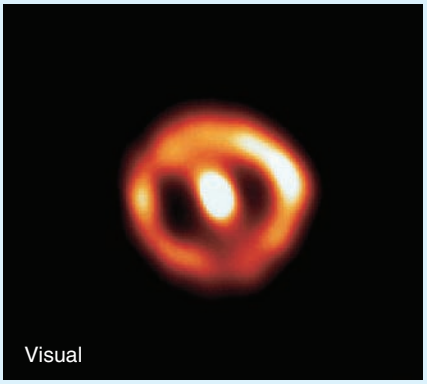
Cause and Effect

Why is the principle of cause and effect so important to scientists?

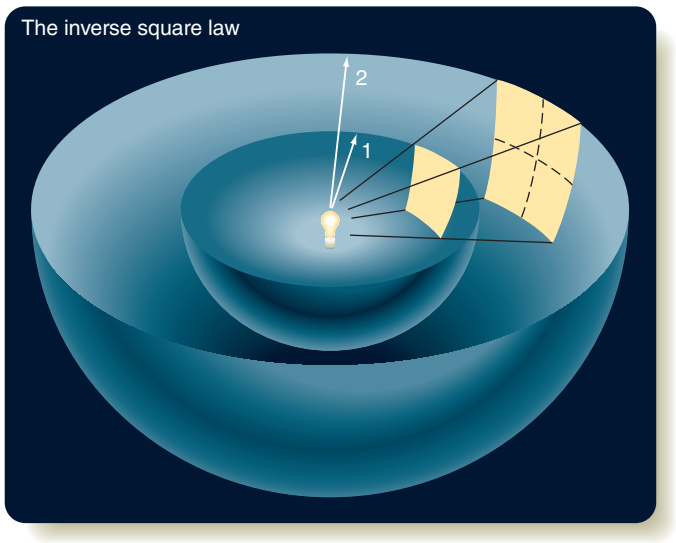
One of the most often used, and least often stated, principles of science is cause and effect. Modern scientists all believe that events have causes, but ancient philosophers such as Aristotle argued that objects moved because of tendencies. They said that earth and water, and objects made mostly of earth and water, had a natural tendency to move toward the center of the Universe. This natural motion had no cause but was inherent in the nature of the objects. Newton's second law of motion ($F = ma$) was the first clear statement of the principle of cause and effect. If an object (of mass m) changes its motion by a certain amount (a in the equation), then it must be acted on by a force of a certain size (F in the equation). That effect (a) must be the result of a cause (F).

The principle of cause and effect goes far beyond motion. Scientists have confidence that every effect has a cause. The struggle against disease is an example. Cholera is a horrible disease that can kill its victims in hours. Long ago it was blamed on such things as bad magic or the will of the gods, and only two centuries ago it was blamed on "bad air." When an epidemic of cholera struck England in 1854, Dr. John Snow carefully mapped cases in London showing that the victims had drunk water from a small number of wells contaminated by sewage. In 1876, the German Dr. Robert Koch traced cholera to an even more specific cause when he identified the microscopic bacillus that causes the disease. Step by step, scientists tracked down the cause of cholera.

If the Universe did not depend on cause and effect, then you could never expect to understand how nature works. Newton's second law of motion was arguably the first clear statement that the behavior of the Universe depends on causes.



Cause and effect: Why did this star explode in 1992? There must have been a cause.



▲ **Figure 5-5** As light radiates away from a source, it spreads out and becomes less intense. Here the amount of light falling on 1 square meter on the inner sphere must cover 4 square meters on a sphere twice as big. This shows how the intensity of light is inversely proportional to the square of the distance. Newton guessed correctly that the force of gravity is inversely proportional to the square of the distance from a mass.

Moon should be about 60^2 (that is, 3600) times weaker than at Earth's surface. The acceleration as a result of Earth's gravity is 9.8 m/s^2 at Earth's surface, so Newton estimated it should be about 0.0027 m/s^2 at the distance of the Moon.

Now, Newton wondered, was this enough acceleration to keep the Moon in orbit? He knew the Moon's distance and its orbital period, so he could calculate the actual acceleration needed to keep it in its curved path. The answer turned out to be 0.0027 m/s^2, as his inverse-square-law calculations predicted. Thus, Newton became certain that the Moon is held in its orbit by gravity, and gravity obeys the inverse square law.

Newton's third law says that forces always occur in pairs, so if Earth pulls on the Moon, then the Moon must pull on Earth. This is called mutual gravitation and is a general property of the Universe. The Sun, the planets, and all their moons must also attract each other by mutual gravitation. In fact, every particle with mass in the Universe must attract every other particle, which is why Newtonian gravity is often called universal mutual gravitation.

Clearly the force of gravity depends on mass. Your body is made of matter, and you have your own personal gravitational field, but your gravity is weak and does not cause personal satellites to orbit around you. Larger masses have stronger gravity. From an analysis of the third law of motion, Newton realized that the mass that resists acceleration in the first law must be identical to the mass causing gravity. Newton performed precise experiments with pendulums and confirmed this equivalence between the mass that resists acceleration and the mass that produces gravity.

From this, combined with the inverse square law, he was able to write the famous formula for the gravitational force between two masses, M and m:

$$F = -\frac{GMm}{r^2}$$

The constant G is the gravitational constant that connects units of mass to units of gravitational force. In the equation, r is the distance between the masses. The negative sign means that the force is attractive, pulling the masses together and making r decrease. In plain language, Newton's law of gravitation states: The force of gravitational attraction between two masses, M and m, is proportional to the product of the masses and inversely proportional to the square of the distance between them.

Newton's description of gravity was a difficult idea for scientists of his time to accept because it involved the puzzling notion of action at a distance. In other words, Earth and Moon somehow exert forces on each other even though there is no physical connection between them. Modern scientists conceptualize this by referring to gravity as a **field**. Earth's mass produces a gravitational field throughout space that is directed toward Earth's center. The strength of the field decreases according to the inverse square law. Any particle with mass in that field experiences a force that depends on the mass of the particle and the strength of the field at the particle's location. The force is directed toward the center of the field.

A field is an elegant way to describe gravity, but it still does not say what gravity is and why there is a field. Later in this chapter, when you learn about Einstein's theory of curved spacetime, you may get a better idea of what gravity really is.

Practicing SCIENCE

***What do the words* universal *and* mutual *mean in the phrase* universal mutual gravitation?** Scientists often work by making step-by-step logical arguments involving careful definitions of words, and this is a good example.

Newton argued that the force that makes an apple accelerate downward is the same as the force that accelerates the Moon and holds it in its orbit. The third law of motion says that forces always occur in pairs, so if Earth attracts the Moon, then the Moon must attract Earth. That is, gravitation is mutual between any two objects.

Furthermore, if Earth's gravity attracts the apple and the Moon, then it must attract the Sun, and the third law says that the Sun must attract Earth. But if the Sun attracts Earth, then it must also attract the other planets and even distant stars, which, in turn, must attract the Sun and each other. Step by step, Newton's third law of motion leads logically to the conclusion that gravitation must apply to all masses in the Universe. That is, gravitation must be universal.

5-2 Orbital Motion and Tides

Orbital motion and tides are two different kinds of gravitational phenomena. As you think about the orbital motion of the Moon and planets, you are considering how gravity pulls on an entire object. When you think about tides, you are considering how gravity pulls on different parts of an object. Analyzing these two phenomena will give you a deeper insight into how gravity works.

5-2a Orbits

Newton was the first person to realize that objects in orbit are falling. You can explore Newton's insight by analyzing the motion of objects orbiting Earth. Carefully read **Orbits** and notice three important concepts and six new terms:

1. An object orbiting Earth is actually falling (being accelerated) toward Earth's center. The object continuously misses colliding with Earth because of its lateral ("sideways") orbital velocity. To follow a circular orbit, the object must move at *circular velocity* (V_c). Placed in a circular orbit at the right distance from Earth, it could be an especially useful *geosynchronous satellite.*
2. Notice that it is more accurate to say that objects orbiting each other are actually revolving around their mutual *center of mass.*
3. Finally, notice the difference between *closed orbits* and *open orbits.* If you want to leave Earth forever, you need to accelerate your spaceship at least until it is moving at *escape velocity* (V_e), so it will follow an *open orbit* and never return.

5-2b Orbital Velocity

To successfully ride a rocket into orbit, you first need to answer a critical question: "How fast must I go to stay in orbit?" An object's circular velocity is the lateral velocity it must have to remain in a circular orbit. If you assume that the mass of your spaceship is small compared with the mass of Earth, then the circular velocity is:

KEY EQUATION 5-1a Circular Velocity

$$V_c = \sqrt{\frac{GM}{r}}$$

In this formula, M is the mass of the central body (Earth in this case) in kilograms, r is the radius of the orbit in meters, and G is the gravitational constant, 6.67×10^{-11} m^3/s^2/kg. This simple formula is all you need to calculate how fast an object must travel to stay in a circular orbit.

Example: How fast does the Moon travel in its orbit? (Assume that the Moon's orbit is perfectly circular even though it is actually slightly elliptical.) Earth's mass is 5.97×10^{24} kg, and the radius of the Moon's orbit is 3.84×10^{8} m. ***Solution:*** The Moon's orbital velocity is:

$$V_c = \sqrt{\frac{6.67 \times 10^{-11} \times 5.97 \times 10^{24}}{3.84 \times 10^{8}}} = \sqrt{\frac{39.8 \times 10^{13}}{3.84 \times 10^{8}}}$$

$$= \sqrt{1.04 \times 10^{6}} = 1020 \text{ m/s} = 1.02 \text{ km/s}$$

This calculation shows that the Moon travels 1.02 km along its orbit each second. That is the circular velocity at the average distance of the Moon.

A satellite just above Earth's atmosphere is only about 200 km above Earth's surface, or 6570 km from Earth's center; so Earth's gravity is much stronger than at the Moon's position and the satellite must travel much faster than the Moon to stay in a circular orbit. You can use the preceding formula to find that the circular velocity for a low orbit 200 km above Earth's surface—just above the atmosphere—is about 7790 m/s, or 7.9 km/s. This is about 17,400 miles per hour, which shows why putting satellites into Earth orbit takes such large rockets. Not only must the rocket lift the satellite above Earth's atmosphere, but the rocket's trajectory must also then curve over and accelerate the satellite horizontally to reach this circular velocity.

A **Common Misconception** holds that there is no gravity in space. You can see that space is filled with gravitational forces from Earth, the Sun, and all other objects in the Universe. An astronaut who appears weightless in space is actually falling along a path at the urging of the combined gravitational fields of the rest of the Universe. Just above Earth's atmosphere, the orbital motion of the astronaut is almost completely due to Earth's gravity.

5-2c Calculating Escape Velocity

If you launch a rocket upward, it will consume its fuel in a few moments and reach its maximum speed. From that point on, it will coast upward. How fast must a rocket travel to coast away from Earth and escape? Of course, no matter how far it travels, it can never escape from Earth's gravity. The effects of Earth's gravity (and the gravity of all other objects) extend to infinity. It is possible, however, for a rocket to travel so fast initially that gravity can never slow it to a stop. Then the rocket could leave Earth permanently.

Escape velocity is the velocity required to escape an astronomical body. Here you are interested in escaping from the surface of Earth; in later chapters you will consider the escape velocity from other planets, the Sun, stars, galaxies, and even black holes.

Escape velocity, V_e, is given by a simple formula:

KEY EQUATION 5-1b Escape Velocity

$$V_e = \sqrt{\frac{2GM}{r}}$$

Again, G is the gravitational constant, 6.67×10^{-11} m^3/s^2/kg, M is the mass of the central body in kilograms, and r is its radius in meters. (Notice that this formula is similar to the formula for circular velocity; in fact, the escape velocity is $\sqrt{2}$ times the circular velocity.)

Example: What is the escape velocity from Earth? ***Solution:*** You can find the escape velocity from Earth by again using its mass, 5.97×10^{24} kg, and the value of Earth's average radius, 6.37×10^{6} m. The escape velocity from Earth's surface is:

$$V_e = \sqrt{\frac{2 \times 6.67 \times 10^{-11} \times 5.97 \times 10^{24}}{6.37 \times 10^{6}}} = \sqrt{\frac{7.96 \times 10^{14}}{6.37 \times 10^{6}}}$$

$$= \sqrt{1.25 \times 10^{8}} = 11{,}200 \text{ m/s} = 11.2 \text{ km/s}$$

This is equal to 11.2 km/s, or about 25,000 mph.

Notice from the formula that the escape velocity from a body depends on both its mass and radius. A massive body might have a low escape velocity if it has a large radius. You will meet such objects when you consider giant stars. On the other hand, a rather low-mass body could have a large escape velocity if it had a small radius, a condition you will encounter when you study black holes.

Once Newton understood gravity and motion, he could do what Kepler had not done; he could explain why the planets obey Kepler's laws of planetary motion.

5-2d Kepler's Laws Revisited

Now that you understand Newton's laws, gravity, and orbital motion, you can look at Kepler's laws of planetary motion in a new and more sophisticated way.

Kepler's first law states that the orbits of the planets are ellipses with the Sun at one focus. In one of his most famous mathematical proofs, Newton showed that if a planet moves in a closed orbit under the influence of an attractive force that follows the inverse square law, then the orbit must be an ellipse. In other words, orbits of planets are ellipses because gravity follows the inverse square law.

Even though Kepler correctly identified the shape of the planets' orbits, he still wondered why the planets keep moving along these orbits, and now you know the answer. They move because there is nothing to slow them down. Newton's first law says

Orbits

1 You can understand orbital motion by thinking of a cannonball moving around Earth in a circular path. Imagine a cannon on a high mountain aimed horizontally as shown at right. A little gunpowder gives the cannonball a low velocity, and it doesn't travel very far before falling to Earth. More gunpowder gives the cannonball a higher velocity, and it travels farther. With enough gunpowder, the cannonball travels so fast it never strikes the ground. Earth's gravity pulls it toward Earth's center, but Earth's surface curves away from it at the same rate it falls. It is in orbit. The velocity needed to stay in a circular orbit is called **circular velocity**. Just above Earth's atmosphere, at an altitude of 200 km, circular velocity is 7790 m/s or about 17,400 miles per hour, and the orbital period is about 88 minutes.

A satellite above Earth's atmosphere feels no friction and will fall around Earth indefinitely.

Earth satellites eventually fall back to Earth if they orbit too low and experience friction with the upper atmosphere.

1a A **geosynchronous satellite** orbits eastward with the rotation of Earth and remains above a fixed spot on the equator, which is ideal for communications and weather satellites.

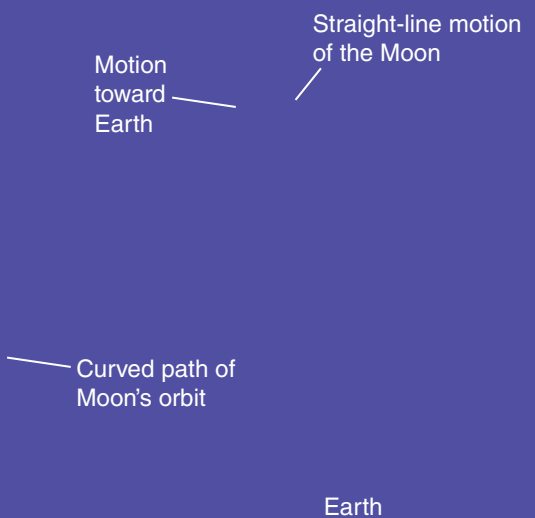
A Geosynchronous Satellite

At a distance of 42,230 km (26,240 mi) from Earth's center, a satellite orbits with a period of 24 hours.

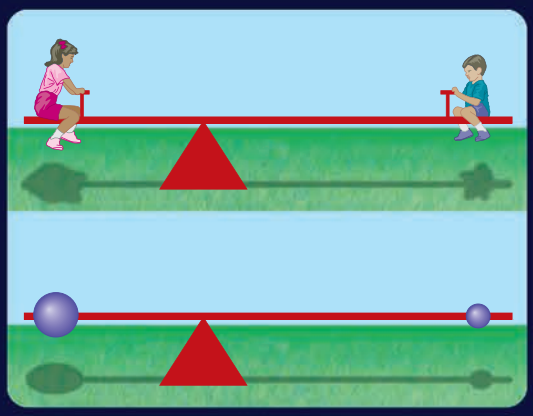
The satellite orbits eastward, and Earth rotates eastward under the moving satellite.

The satellite remains fixed above a spot on Earth's equator.

1b According to Newton's first law of motion, the Moon should follow a straight line and leave Earth forever. Because it follows a curve, Newton knew that some force must continuously accelerate it toward Earth, gravity. Every second the Moon moves 1020 m (3350 ft) eastward and falls about 1.4 mm (1/18 in.) toward Earth. The combination of these motions produces the Moon's curved orbit. The Moon is falling *all* the time.

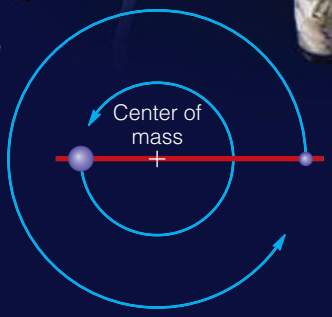


1c Astronauts in orbit around Earth feel weightless, but they are not—to use a term from old science fiction movies—"beyond Earth's gravity." Like the Moon, the astronauts are accelerated toward Earth by Earth's gravity, but they travel fast enough along their orbits that they continually "miss the Earth." They are literally falling around Earth. Inside or outside a spacecraft, astronauts feel weightless because they and their spacecraft are falling at the same rate. Rather than saying they are weightless, you should more accurately say they are in free fall.

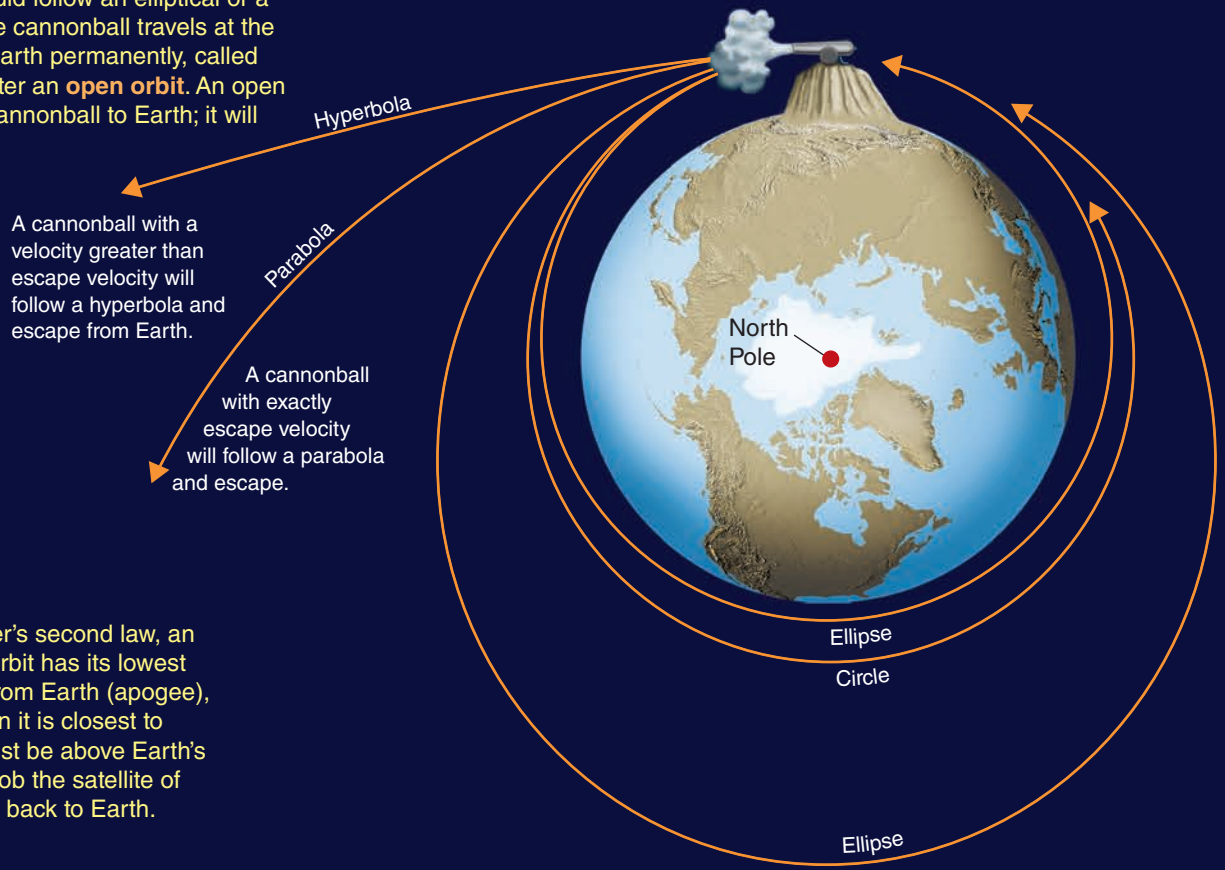


2 To be accurate you should not say that an object orbits Earth. Rather, the two objects orbit each other. Gravitation is mutual, and if Earth pulls on the Moon, the Moon pulls on Earth. The two bodies revolve around their common **center of mass**, the balance point of the system.

2a Two bodies of different mass balance at their center of mass, which is located closer to the more massive object. As the two objects orbit each other, they revolve around their common center of mass as shown at right. The center of mass of the Earth–Moon system lies only 4670 km (2900 mi) from the center of Earth—inside Earth. As the Moon orbits the center of mass on one side, Earth swings around the center of mass on the opposite side.



3 Closed orbits are repeating cycles. The Moon and artificial satellites orbit Earth in closed orbits. Below, the cannonball could follow an elliptical or a circular **closed orbit**. If the cannonball travels at the velocity needed to leave Earth permanently, called **escape velocity**, it will enter an **open orbit**. An open orbit does not return the cannonball to Earth; it will escape.


Hyperbola
A cannonball with a velocity greater than escape velocity will follow a hyperbola and escape from Earth.
Parabola
A cannonball with exactly escape velocity will follow a parabola and escape.
North Pole
Ellipse
Circle
Ellipse

3a As described by Kepler's second law, an object in an elliptical orbit has its lowest velocity when it is farthest from Earth (apogee), and its highest velocity when it is closest to Earth (perigee). Perigee must be above Earth's atmosphere, or friction will rob the satellite of energy and it will quickly fall back to Earth.

that a body in motion stays in motion unless acted on by some force. In the absence of friction, the planets must continue to move.

Kepler's second law states that a planet moves faster when it is near the Sun and slower when it is farther away. Once again, Newton's discoveries explain why. Imagine you are in an elliptical orbit around the Sun. As you move around the most distant part of the ellipse, aphelion, you begin to move back closer to the Sun, and the Sun's gravity pulls you slightly forward in your orbit. You pick up speed as you fall toward the Sun, so, of course, you go faster as you approach the Sun. As you move around the closest point to the Sun, perihelion, you begin to move away from the Sun, and the Sun's gravity pulls slightly backward on you, slowing you down as you recede from the Sun. If you were in a circular orbit, the Sun's gravity would always pull perpendicular to your motion, and you would not speed up or slow down. Kepler's second law makes sense when you analyze it in terms of forces and motions.

There is a more elegant and profound way to think about Kepler's second law. Previously you learned about Galileo's insight that a body moving on a frictionless surface continues to move in a straight line until it is acted on by some force; that is, the object has momentum. In a similar way, an object rotating on a frictionless surface will continue rotating until something acts to speed up or slow down its rotation. Such an object has **angular momentum,** a combination of the object's mass with its speed of rotation or revolution. A planet circling the Sun in an orbit has a given amount of angular momentum, and, with no outside influences to alter its motion, its angular momentum must remain constant; physicists say that angular momentum is "conserved." Mathematically, a planet's angular momentum around the Sun is the product of its mass, velocity, and distance from the Sun. This provides another way to understand why a planet must speed up as it comes closer to the Sun along an elliptical orbit. Because its angular momentum is conserved, as its distance from the Sun decreases its velocity must increase, and as its distance from the Sun increases its velocity must decrease.

The conservation of angular momentum is actually a common human experience. Skaters spinning slowly can draw their arms and legs closer to their axis of rotation and, through conservation of angular momentum, spin faster (Figure 5-6). To slow their rotation, they can extend their arms again. Similarly, divers can spin rapidly in the tuck position and then slow their rotation by stretching into the extended position.

Kepler's third law states that a planet's orbital period depends on its distance from the Sun. The fundamental reason for that law lies in the fact that the **energy** of the planet's motion is determined by its distance from the Sun. Physicists define energy as "the ability to do work," but you might paraphrase that as "the ability to produce a change." A planet moving along its orbit, a cement truck rolling down the highway, and a golf ball zipping down the fairway all have the ability to produce changes. Imagine colliding with any of those objects! [Amounts of energy are expressed in **joules (J)** in the metric system.]

▲ **Figure 5-6** Skaters demonstrate conservation of angular momentum when they spin faster by drawing their arms and legs closer to their axis of rotation.

A planet orbiting the Sun has a specific amount of energy that is the sum of the energy of motion, called *kinetic energy*, plus energy involved in the gravitational attraction between the planet and the Sun, called *potential energy*. Because both types of energy depend on the planet's distance from the Sun, there must be a fixed relationship between the rate at which a planet moves around its orbit and the size of the orbit—between its orbital period, P, and the orbit's semimajor axis, a. The relation between those two kinds of energy underlies Kepler's and Newton's laws. You can even derive Kepler's third law from Newton's laws of motion, as shown in the next section.

5-2e Newton's Version of Kepler's Third Law

The equation for circular velocity is actually a version of Kepler's third law, as you can prove with three lines of simple algebra. The result is one of the most useful formulas in astronomy.

The equation for circular velocity, as you have seen, is:

$$V_c = \sqrt{\frac{GM}{r}}$$

The circular orbital velocity of a planet is simply the circumference of its orbit divided by the orbital period:

$$V = \frac{2\pi r}{P}$$

If you substitute this for V in the first equation and solve for P^2, you get:

KEY EQUATION 5-2 Newton's Version of Kepler's 3rd Law

$$P^2 = \left(\frac{4\pi^2}{GM}\right) r^3$$

Here M is the total mass of the two-body system in kilograms. For a planet orbiting the Sun, you can use just the mass of the Sun as a good approximation for M because the mass of the planet is negligible compared to the mass of the Sun, so adding in the planet's mass makes only a tiny difference. (In a later chapter, you will apply this formula to two stars orbiting each other, and then you will need to include both masses in the sum.) For a circular orbit, the semimajor axis, *a,* equals the radius of the circle.

Example: What is the orbital period of the Moon around Earth? ***Solution:*** The Sun's mass is 1.989×10^{30} kg, and the average distance of Earth from the Sun is 1.496×10^{11} m. The gravitational constant G is 6.674×10^{-11} $m^3/s^2/kg$. (Note that for this example the input quantities are given to 4 digits of precision.) Assume that the mass of Earth is negligible compared with the mass of the Sun.

$$P = \sqrt{\frac{4 \times \pi^2 \times (1.496 \times 10^{11})^3}{6.674 \times 10^{-11} \times 1.989 \times 10^{30}}} = \sqrt{\frac{1.322 \times 10^{35}}{1.327 \times 10^{20}}}$$

$$= \sqrt{9.957 \times 10^{14}} = 3.155 \times 10^7 \text{ s}$$

The result is 3.155×10^7 seconds = 365¼ days.

As you can see, this formula is a general version of Kepler's third law, $P^2 = a^3$, and you can use it for any object orbiting any other object. In Kepler's version, you used astronomical units (AU) for distance and years for time, but in Newton's version of the formula, you need to use units of meters, seconds, and kilograms. G is the gravitational constant, defined previously.

This is a powerful formula. If, for example, you observe a moon orbiting a planet and you can measure the size of that moon's orbit, *r,* and its orbital period, *P,* you can use this formula to solve for *M,* the total mass of the planet plus the moon. It is important for you to realize that *there is no other way to precisely measure masses of objects in the Universe than by using this formula*. In later chapters, you will see this formula used over and over to find the masses of stars, galaxies, and planets.

This discussion is a good illustration of the power of Newton's work. By carefully defining motion and gravity and by giving them mathematical expression, Newton was able to derive new truths, among them his version of Kepler's third law. His work finished the transformation of what were once considered the mysterious wanderings of the planets into understandable motions that follow simple rules. In fact, his discovery of gravity explained something else that had mystified philosophers for millennia: the ebb and flow of ocean tides.

5-2f Tides and Tidal Forces

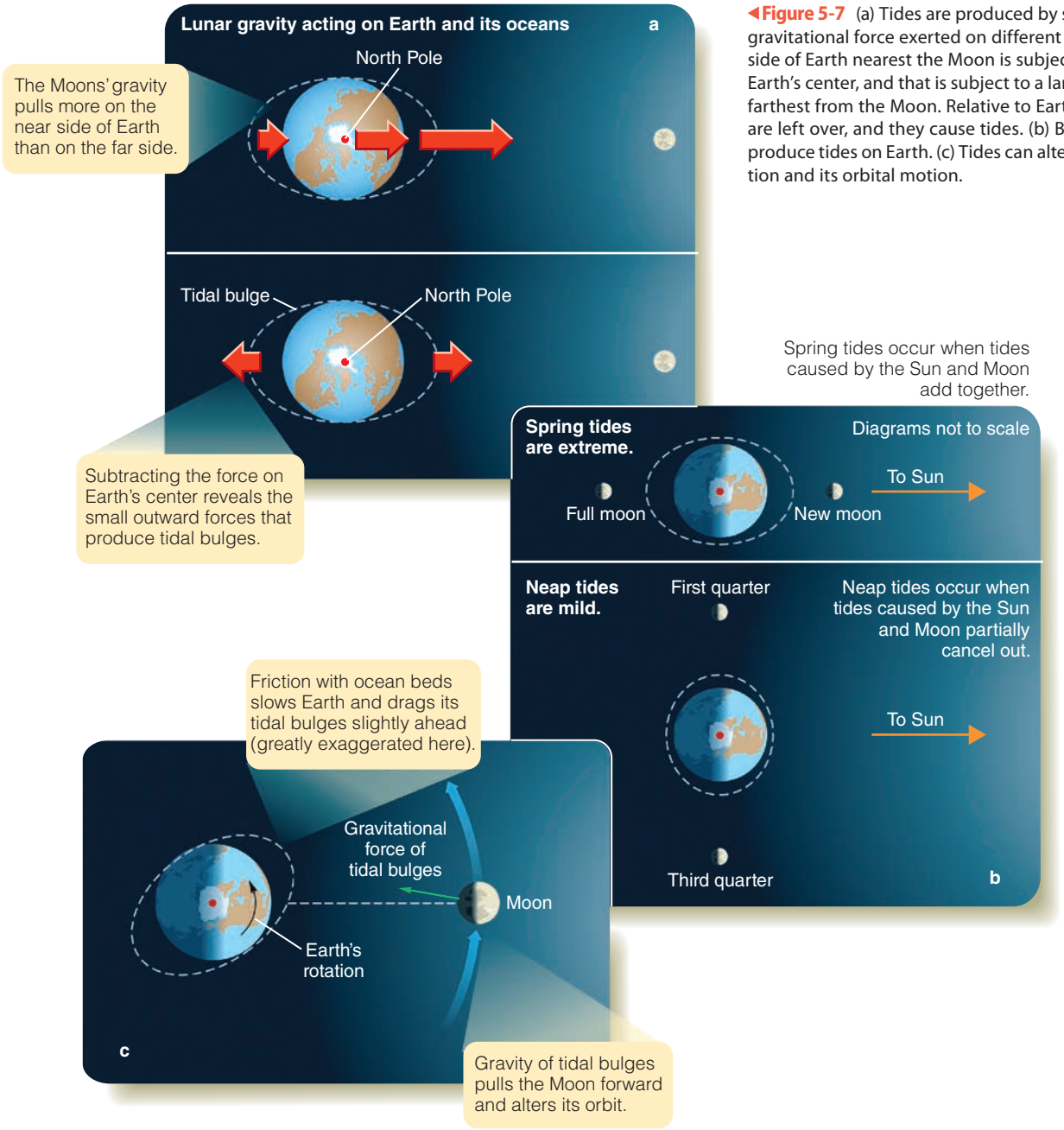
Newton understood that gravity is mutual—Earth attracts the Moon, and the Moon attracts Earth—and that means the Moon's gravity can explain the ocean tides.

Tides are caused by small differences in gravitational forces. For example, Earth's gravity attracts your body downward with a force equal to your weight. The Moon is less massive and more distant, so it attracts your body with a force that is a tiny percentage of your weight. You don't notice that little force, but Earth's oceans respond visibly.

The side of Earth that faces the Moon is about 6400 km (4000 mi) closer to the Moon than is the center of Earth. Consequently, the Moon's gravity, small though it is at the distance of Earth, is just a bit stronger on the near side of Earth than on the center. It pulls on the oceans on the near side of Earth a bit more strongly than on Earth's center, and the oceans respond by flowing to make a bulge of water on the side of Earth facing the Moon. There is also a bulge on the side of Earth that faces away from the Moon because the Moon pulls more strongly on Earth's center than on its far side. Thus, on the far side of Earth the Moon pulls Earth away from the oceans, which flow into a second bulge, this one pointing away from the Moon, as shown in **Figure 5-7a**. The ocean tides are caused by the accelerations Earth and its oceans feel as they orbit around the Earth–Moon center of mass.

A **Common Misconception** holds that the Moon's effect on tides means that the Moon has an affinity for water—including the water in your body—and, according to some people, that's how the Moon affects you. That's not true. If the Moon's gravity affected only water, then there would be only one tidal bulge, the one facing the Moon. As you know, the Moon's gravity acts on all of Earth, the rock as well as the water, and that produces the tidal bulge in the oceans on the far side of Earth. In fact, small tidal bulges occur in the rocky bulk of Earth because it is deformed by the Moon's gravity. Although you do not notice it, as Earth rotates the landscape rises and falls by a few centimeters with the tides. The Moon has no special affinity for water, and, because your body is so much smaller than Earth, any tides the Moon raises in your body are immeasurably small. Ocean tides are large because oceans are large.

You can see obvious evidence of tides if you watch the ocean shore for a few hours. The tidal bulges remain fixed in position with respect to the Moon as Earth rotates. The turning Earth carries you and your beach into a tidal bulge, the ocean water deepens, and you see the tide crawling up the sand. The tide does not really "come in"; it's more accurate to say you move into the tidal bulge. Later, when Earth's rotation carries you out of the bulge, the ocean becomes shallower, and the tide falls. The tides rise and fall twice a day on a normal coastline because there are two bulges on opposite sides of Earth.


Lunar gravity acting on Earth and its oceans
a
North Pole
The Moons' gravity pulls more on the near side of Earth than on the far side.
Tidal bulge
North Pole
Subtracting the force on Earth's center reveals the small outward forces that produce tidal bulges.
Spring tides occur when tides caused by the Sun and Moon add together.
Spring tides are extreme.
Diagrams not to scale
To Sun
Full moon
New moon
Neap tides are mild.
First quarter
Neap tides occur when tides caused by the Sun and Moon partially cancel out.
To Sun
Third quarter
b
Friction with ocean beds slows Earth and drags its tidal bulges slightly ahead (greatly exaggerated here).
Gravitational force of tidal bulges
Moon
Earth's rotation
c
Gravity of tidal bulges pulls the Moon forward and alters its orbit.

◀**Figure 5-7** (a) Tides are produced by small differences in the gravitational force exerted on different parts of an object. The side of Earth nearest the Moon is subject to a larger force than Earth's center, and that is subject to a larger force than the side farthest from the Moon. Relative to Earth's center, small forces are left over, and they cause tides. (b) Both the Moon and Sun produce tides on Earth. (c) Tides can alter both an object's rotation and its orbital motion.

The tidal cycle at any given location can be quite complex because it is affected by the latitude of the site, shape of the shoreline, wind strength, and so on. For example, tides in the Bay of Fundy (New Brunswick, Canada) occur twice a day and can exceed 40 feet (12 m). In contrast, the northern coast of the Gulf of Mexico has only one tidal cycle a day of roughly 1 foot (30 cm).

Gravity is universal, so the Sun also produces tides on Earth. The Sun is 27 million times more massive than the Moon, but it lies almost 400 times farther from Earth. Tides on Earth caused by the Sun are less than half as high as those caused by the Moon. Twice a month, at new moon and at full moon, the Moon and Sun produce tidal bulges that add together and produce extreme tidal changes: At those moon phases, high tides are exceptionally high, and low tides are exceptionally low. Such tides are called **spring tides**. Here the word *spring* does not refer to the season of the year but to the rising up of water. At first- and third-quarter moons, the Sun and Moon pull at right angles to each other, and the tides caused by the Sun partly cancel out the tides caused by the Moon. These less extreme tides are called **neap tides**. The word *neap* comes from an Old English word, *nep,* that meant something like *weak*. Spring tides and neap tides are illustrated in **Figure 5-7b**.

Galileo tried to understand tides, but it was not until Newton described gravity that astronomers could analyze tidal

forces and recognize their surprising effects. For example, the moving water in tidal bulges experiences friction with the ocean beds and resistance as it rises onto continents. That friction slows Earth's rotation and currently makes the length of a day grow by 0.0023 second per century. Thin layers of silt laid down millions of years ago where rivers emptied into oceans contain a record of tidal cycles as well as daily, monthly, and annual cycles. Those data confirm that only 620 million years ago Earth's day was less than 22 hours long.

Tidal forces can also affect orbital motion. Earth rotates eastward, and friction with the ocean beds drags the tidal bulges slightly eastward out of a direct Earth–Moon line. These tidal bulges are massive, and their gravitational field pulls the Moon forward in its orbit, as shown in **Figure 5-7c**. As a result, the Moon's orbit is growing larger by about 3.8 cm a year, an effect that astronomers can measure by bouncing laser beams off reflectors left on the lunar surface by the Apollo astronauts.

Earth's gravitation exerts tidal forces on the Moon, and, although there are no bodies of water on the Moon, friction within the flexing rock has slowed the Moon's rotation to the point that it now keeps the same face toward Earth.

Tides are much more than just the cause of oceans rising and falling in daily and monthly rhythms. In later chapters, you will see how tides can pull gas away from stars to feed black holes, rip galaxies apart, and melt the interiors of small moons orbiting massive planets. Tidal forces produce some of the most surprising and impressive processes in the Universe.

5-2g Astronomy After Newton

Newton published his work in 1687 in a book titled, in Latin, *Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy),* now known simply as the *Principia* (pronounced *prin-KIP-ee-uh*). It is one of the most important books ever written. The *Principia* changed astronomy, science, and the way people think about nature.

The *Principia* changed astronomy by ushering in a new age. No longer did people have to appeal to whims of the gods to explain things in the heavens. No longer did they speculate on why the planets wander across the sky. After the *Principia* was published, physicists and astronomers understood that the motions of celestial bodies are governed by simple, universal rules that describe the motions of everything from orbiting planets to falling apples. Suddenly the Universe was understandable in simple terms, and astronomers could accurately predict future planetary motions (**How Do We Know? 5-2**).

The *Principia* also changed science in general. The works of Copernicus and Kepler had been mathematical, but no book before the *Principia* had so clearly demonstrated the power of mathematics as a language of precision. Newton's arguments in his book were such powerful illustrations of the quantitative study of nature that scientists around the world adopted mathematics as their most powerful tool.

Also, the *Principia* changed the way people thought about nature. Newton showed that the rules that govern the Universe are simple. Particles move according to just three laws of motion and attract each other with a force called *gravity.* These motions are predictable, and that makes the Universe seem like a vast machine, but one with operations based on a few simple rules. The Universe is complex only in that it contains a vast number of particles. In Newton's view, if he knew the location and motion of every particle in the Universe, he could—in principle—derive the past and future of the Universe in every detail. This idea of mechanical determinism has been substantially modified by the work of physicists during the 20th century, but it dominated science for more than two centuries. During those years, scientists thought of nature primarily as a beautiful clockwork that would be perfectly predictable if they knew how all the gears meshed.

Most of all, Newton's work broke the last bonds between science and formal philosophy. Newton did not speculate on the good or evil of gravity. Not more than a hundred years before, scientists would have argued over the "meaning" of gravity. Newton didn't care for these debates. He wrote, "It is enough that gravity exists and suffices to explain the phenomena of the heavens."

Newton's laws were foundations of astronomy and physics for two centuries. Then, early in the 20th century, a physicist named Albert Einstein proposed a new way to describe gravity. The new theory did not replace Newton's laws but rather showed that they were only approximately correct and could be seriously in error under certain special circumstances. Einstein's theories further extended the scientific understanding of the nature of gravity. Just as Newton stood on the shoulders of Galileo, Einstein stood on the shoulders of Newton.

Practicing SCIENCE

How do Newton's laws of motion and gravity explain the orbital motion of the Moon? What scientific argument—chain of evidence and logical statements—can you use, as scientists have done many times since Newton's day, to verify that the orbit of the Moon shows the operation of Newton's laws of motion and gravity?

If Earth and the Moon did not attract each other, the Moon would move in a straight line in accord with Newton's first law of motion and vanish into deep space. Instead, gravity pulls the Moon toward Earth's center, and the Moon accelerates toward Earth. This acceleration is just enough to pull the Moon away from its straight-line motion and cause it to follow a curve around Earth.

In fact, it is correct to say that the Moon is falling, but because of its lateral motion it continuously misses Earth. Every orbiting object is falling toward the center of its orbit but is also moving laterally fast enough to compensate for the inward motion, and it follows a curved orbit.

Testing a Hypothesis by Prediction

How are the predictions of a hypothesis useful in science?

Scientific hypotheses face in two directions. They look back into the past and explain phenomena previously observed. For example, Newton's laws of motion and gravity explained observations of the movements of the planets made over many centuries. But hypotheses also look forward, making predictions about what you should find as you explore further. For example, Newton's laws allowed astronomers to calculate the orbits of comets, predict their return, and eventually understand their origin.

Scientific predictions are important in two ways. First, if a prediction of a hypothesis is confirmed, scientists gain confidence that the hypothesis is a true description of nature. But predictions are important for a second reason. They can point the way to unexplored avenues of knowledge.

Particle physics is a field in which predictions have played a key role in directing research. In the early 1970s, physicists proposed a hypothesis, later named the *Standard Model*, regarding particles inside atoms and the forces between them. This hypothesis explained what scientists had already observed in experiments, but it also predicted the existence of particles that had not yet been observed. To test the hypothesis, scientists focused their efforts on building more and more powerful particle accelerators in the hopes of detecting the predicted particles.

A number of these particles have since been discovered, and they do match the characteristics predicted by the Standard Model, further confirming the hypothesis. Existence of the Higgs boson, a fundamental particle predicted by the Standard Model, was confirmed in 2013. There are still some predictions of the Standard Model remaining to be checked, and this scientific detective story is ongoing.

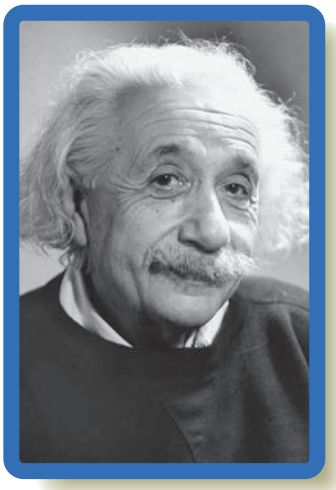
You learned in the previous chapter that a hypothesis that has passed many tests and has wide predictive value can "graduate" to being considered a theory, and if it is considered fundamental enough, it may be called a "law." Newton's laws of motion are at the end of the journey toward powerful reliability from their beginning as tentative hypotheses. The particle physics Standard Model is pretty far along in its version of the same journey. As you read about any scientific hypothesis, think about both what it can explain that has been observed already and what it can predict that can be observed in future.

Brookhaven National Laboratory

Physicists build huge accelerators to search for subatomic particles predicted by their hypotheses.

5-3 Einstein and Relativity

In the early years of the last century, Albert Einstein (1879–1955; **Figure 5-8**) began thinking about how motion and gravity are related. He soon gained international fame by showing that Newton's laws of motion and gravity were only partially correct. The revised theory became known as the theory of relativity. As you will see, there are really two theories of relativity.



Fred Stein Archive/Archive Photos/Getty Images

◀ **Figure 5-8** Einstein has become a symbol of the brilliant scientist. His fame began when he was a young man and thought deeply about the nature of motion. That led him to revolutionary insights into the meaning of space and time and a new understanding of gravity.

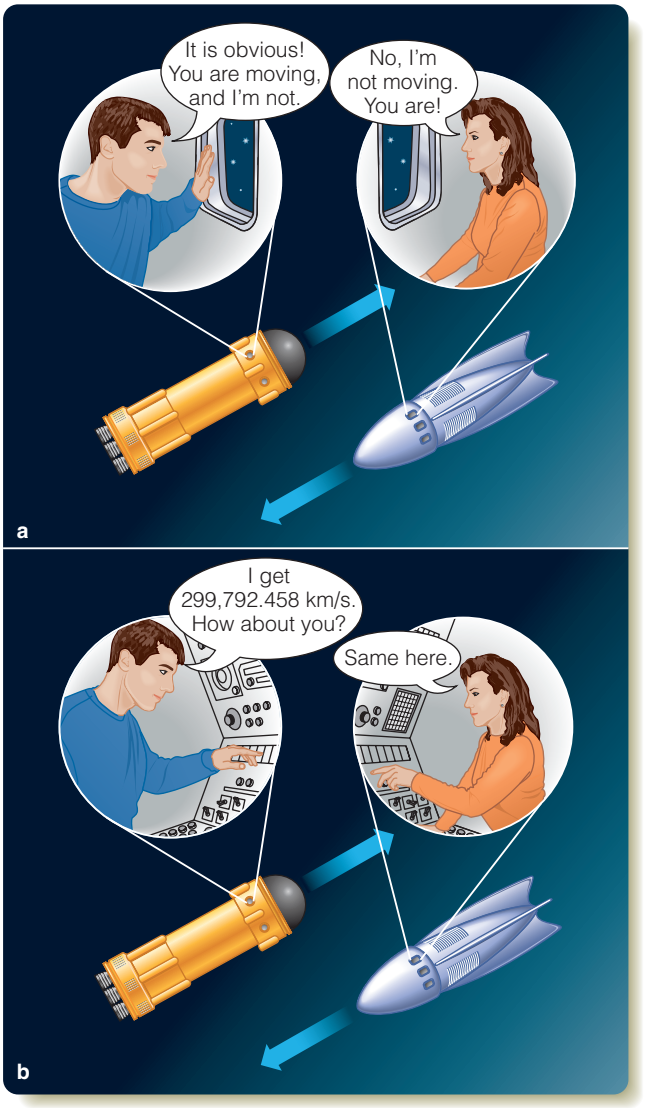
5-3a Special Relativity

Einstein began by thinking about how moving observers see events around them. His analysis led him to the first postulate of relativity, also known as the principle of relativity:

> **First postulate** (the principle of relativity): Observers can never detect their *uniform* motion except relative to other objects.

You may have had experiences that illuminate the first postulate while sitting on a train in a station. You suddenly notice that the train on the next track has begun to creep out of the station. However, after several moments you realize that it is your own train that is moving and that the other train is still motionless on its track. You can't tell which train is moving until you look at external objects such as the station platform.

Consider a second example. Suppose you are floating in a spaceship in interstellar space, and another spaceship comes coasting by (**Figure 5-9a**). You might conclude that it is moving and you are not, but someone in the other ship might be equally sure that you are moving and it is not. Of course, you could just look out a window and compare the motion of your spaceship with a nearby star, but that just expands the problem. Which is



▲ **Figure 5-9** (a) The principle of relativity says that observers can never detect their uniform motion, except relative to other objects. Neither of these travelers can decide who is moving and who is not. (b) If the speed of light depended on the motion of the observer through space, then these travelers could perform measurements inside their spaceships to discover who was moving. If the principle of relativity is correct, then the speed of light must be a constant when measured by any observer.

moving, your spaceship or the star? The principle of relativity says that there is no experiment you can perform inside your ship to decide which ship is moving and which is not. This means that all motion is relative.

Because no internal experiment can detect either spaceship's absolute motion through space, the laws of physics must have the same form inside both ships. Otherwise, experiments would produce different results in the two ships, and you could decide which was moving. Thus, a more general way of stating Einstein's first postulate refers to the laws of physics:

> **First postulate** (more sophisticated version): The laws of physics are the same for all observers, no matter what their motion, so long as they are not *accelerated*.

The word *accelerated* is important. If either spaceship were to fire its rockets, then its velocity would change. The crew of that ship would know it because they would feel the acceleration pressing them into their couches. Accelerated motion, therefore, is different; the pilots of the spaceships can always tell which ship is accelerating and which is not. The postulates of relativity discussed here apply only to the special case of observers in *uniform* motion, which means *unaccelerated* motion. That is why the theory is called the **special theory of relativity**.

The first postulate led Einstein to the conclusion that the speed of light must be constant for all observers. No matter how you are moving, your measurement of the speed of light has to give the same result (Figure 5-9b). This became the second postulate of special relativity:

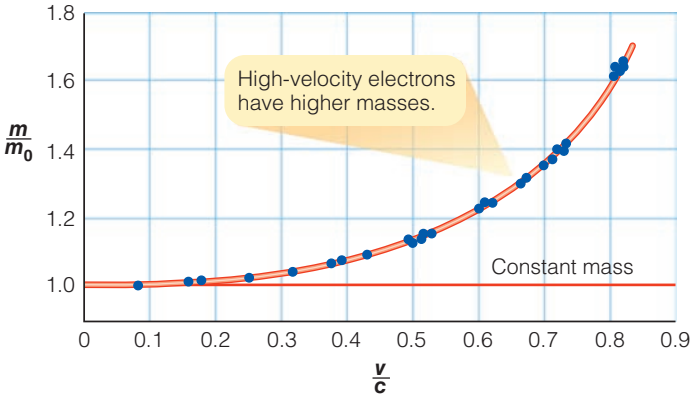
> **Second postulate:** The speed of light in a vacuum is constant and will have the same value for all observers independent of their motion relative to the light source.

You can see that this is required by the first postulate; if the speed of light were not constant, then the pilots of the spaceships could measure the speed of light inside their spaceships and decide who was moving. (Note the phrase "in a vacuum." Light slows down when it passes through a medium; the speed of light in water is one-third less than the speed of light in a vacuum. It is the speed of light in a vacuum, not affected by passing through any medium, to which the first postulate refers.)

Once Einstein had thought through the basic postulates of relativity (Table 5-2), he was led to some startling discoveries. Newton's laws of motion and gravity work well in situations where distances are small and velocities were low. But when Einstein began to think about very large distances or very high velocities, he realized that Newton's laws were no longer always adequate to describe what happens. Instead, the postulates led Einstein to derive a more accurate description of nature that is now known as the special theory of relativity. It predicts some peculiar effects. For example, special relativity predicts that the observed mass of a moving particle depends on its velocity. The higher the velocity, the greater will be the mass of the particle. This effect is not significant at low velocities, but it becomes important as the velocity approaches the speed of light. As strange as

Table 5-2 Postulates of Relativity

I. (Relativity principle) Observers can never detect their *uniform* motion except relative to other objects.
II. The speed of light is constant and will be the same for all observers independent of their motion relative to the light source.
III. (Equivalence principle) Observers cannot distinguish between inertial forces due to acceleration versus uniform gravitational forces.



▲ **Figure 5-10** The observed mass of moving electrons depends on their velocity. As the ratio of their velocity to the velocity of light, *v/c,* increases, the mass of the electrons relative to their mass at rest, m/m_0, increases. Such relativistic effects are quite evident in particle accelerators, which accelerate atomic particles to very high velocities.

that may seem, such increases in mass are reliably observed whenever physicists accelerate particles to high velocities (**Figure 5-10**).

This discovery led to yet another insight. The relativistic equations that describe the energy of a moving particle predict that the energy of a motionless particle is not zero. Rather, its energy at rest is m_0c^2. This is, of course, the famous equation:

KEY EQUATION 5-3 Einstein's Energy-Mass Equivalence

$$E = m_0c^2$$

The constant c is the speed of light, and m_0 is the mass of the particle when it is at rest. This simple formula shows that mass and energy are related, and you will see in later chapters how nature can convert one into the other inside stars.

Example: How much energy is contained in 1 kg of matter? ***Solution:*** The speed of light is 3×10^8 m/s, so your result is 9×10^{16} joules (J), approximately equal to the energy released by a 20-megaton nuclear bomb. Recall that a joule is a unit of energy roughly equivalent to the energy given up when an apple falls from a table to the floor. This simple calculation shows that the energy equivalent of even a small mass is very large.

Other relativistic effects include the slowing of moving clocks and the shrinking of lengths measured in the direction of motion. A detailed discussion of the major consequences of the special theory of relativity is beyond the scope of this book, but you can be confident that these strange effects have been confirmed many times in experiments. Einstein's work is called the special *theory* of relativity because it meets the scientific definition of a *theory:* It is well understood, has been checked many times in many ways, and is widely applicable (look back to How Do We Know? 4-2).

5-3b The General Theory of Relativity

In 1916, Einstein published a more general version of the theory of relativity that dealt with accelerated as well as uniform motion. This **general theory of relativity** contained a new description of gravity.

Einstein began by thinking about observers in accelerated motion. Imagine an observer sitting in a windowless spaceship. Such an observer cannot distinguish between the force of gravity and the inertial forces produced by the acceleration of the spaceship (**Figure 5-11**). This led Einstein to conclude that gravity and acceleration are related, which is a conclusion now known as the equivalence principle:

> **Equivalence principle:** Observers cannot distinguish locally between inertial forces due to acceleration versus uniform gravitational forces due to the presence of a massive body.

This should not surprise you. Previously in this chapter, you read that Newton concluded that the mass that resists acceleration is the same as the mass that exerts gravitational forces and then performed experiments to confirm that principle.

The importance of the general theory of relativity lies in its description of gravity. Einstein concluded that gravity, inertia, and acceleration are all associated with the way space and time are connected as a single entity referred to as space-time. This relation is often referred to as curvature, and a one-line description of general relativity is that it explains a gravitational field as a curved region of space-time:

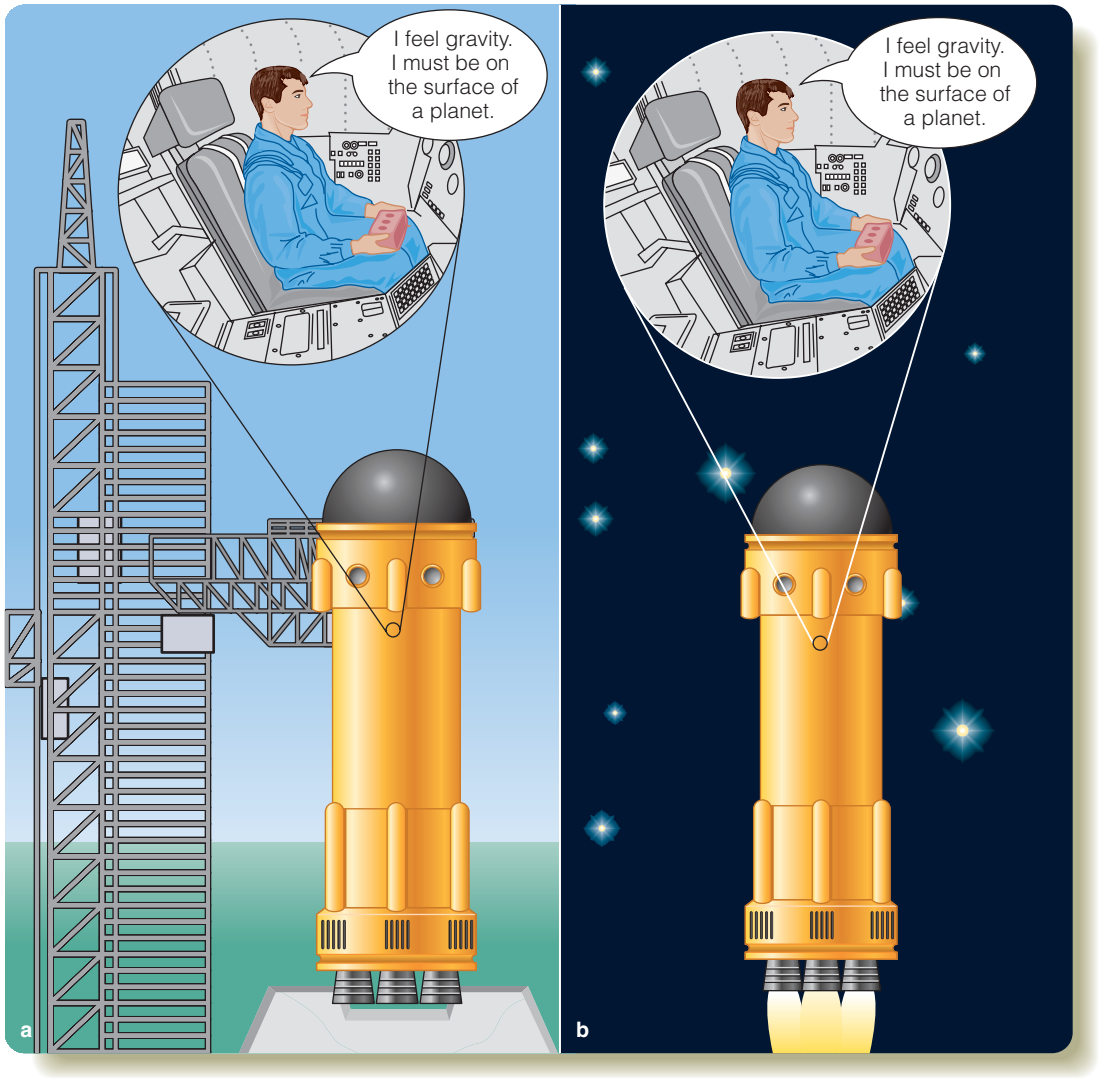
> **Gravity according to general relativity:** Mass tells space-time how to curve, and the curvature of space-time (gravity) tells mass how to accelerate.

Therefore, you feel gravity because Earth's mass causes a curvature of space-time. The mass of your body responds to that curvature by accelerating toward Earth's center, and that presses you downward in your chair. According to general relativity, all masses cause curvature of the space around them, and the larger the mass, the more severe the curvature. That's gravity.

5-3c Confirmation of the Curvature of Space-Time

Einstein's general theory of relativity has been confirmed by a number of experiments, but two are worth mentioning here because they were among the first tests of the theory, and required astronomical observations. One involves Mercury's orbit, and the other involves eclipses of the Sun.

Kepler understood that the orbit of Mercury is elliptical. Later, astronomers discovered that the long axis of Mercury's orbit sweeps around the Sun in a motion that is an example of precession (look back to Figure 2-7). The total observed



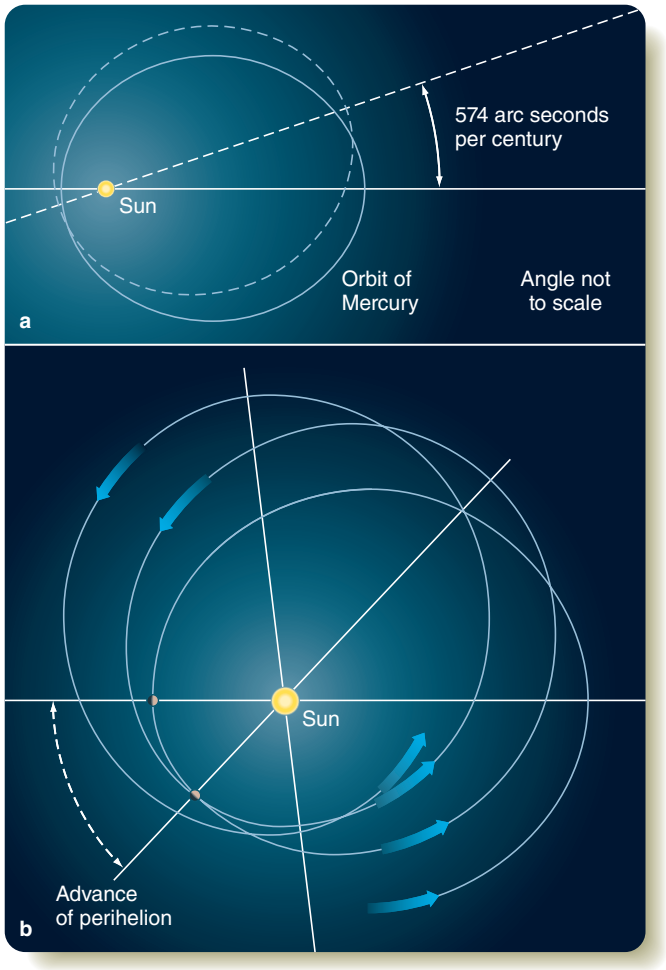
▲ **Figure 5-11** (a) An observer in a closed spaceship on the surface of a planet feels gravity. (b) In space, with the rockets smoothly firing and accelerating the spaceship, the observer feels inertial forces that are equivalent to gravitational forces.

precession is almost 600 arc seconds per century (**Figure 5-12a**). Most of this precession is caused by the gravitation of Venus, Earth, and the other planets. However, when astronomers take all known effects into account and use Newton's description of gravity to account for the gravitational influence of all of the planets, they are left with a small excess. Mercury's orbit is precessing 43 arc seconds per century faster than Newton's laws predict.

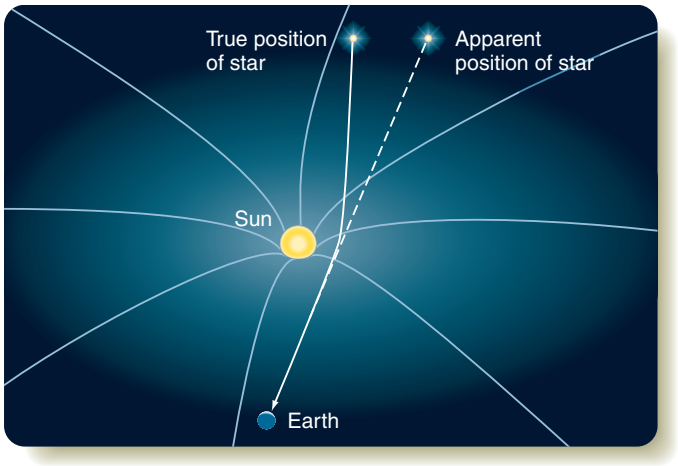
This is a tiny effect. Each time Mercury returns to perihelion—its closest point to the Sun—it is about 29 km (18 mi) past the position predicted by Newton's laws. This is such a small distance compared with the planet's diameter of 4880 km that it could never have been detected had it not been cumulative. Each orbit, Mercury gains only 29 km, but after a century it's ahead by more than 12,000 km—more than twice its own diameter. This tiny effect, called the "advance" of Mercury's orbital perihelion, accumulated from the time of Newton to the time of Einstein into a serious discrepancy in the Newtonian description of the Universe.

The advance of perihelion of Mercury's orbit was one of the first problems to which Einstein applied the principles of general relativity. First he calculated how much the Sun's mass curves space-time in the region of Mercury's orbit, and then he calculated how Mercury moves through the space-time. The theory predicted that the curved space-time should cause Mercury's orbit to advance by 43.03 arc seconds per century, exactly the same as the observed excess to within the measurement uncertainty (**Figure 5-12b**).

When his theory matched observations, Einstein was so excited he could not return to work for three days. He would be even happier with modern studies that have shown that Venus, Earth, and even Icarus, an asteroid that comes close to the Sun,



▲ **Figure 5-12** (a) Mercury's orbit precesses 574 arc seconds per century, 43 arc seconds more than predicted by Newton's laws. (b) Even when you ignore the influences of the other planets, Mercury's orbit is not a perfect ellipse. Curved space-time near the Sun distorts the orbit from an ellipse into a rosette. The advance of Mercury's perihelion is exaggerated by a factor of about 1 million in this figure.



▲ **Figure 5-13** Like a depression in a putting green, the curved space-time near the Sun deflects light from distant stars and makes them appear to lie slightly farther from the Sun than their true positions.

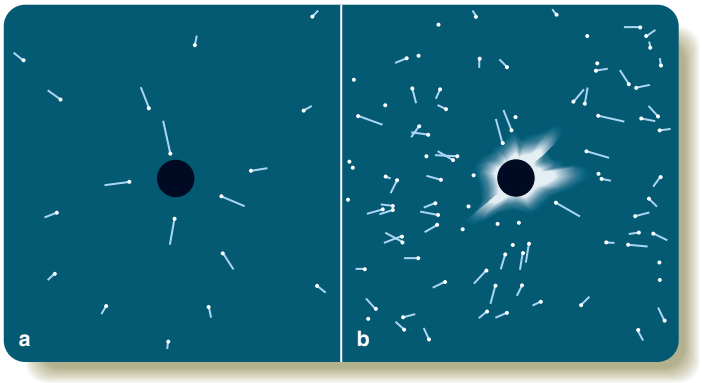
also have orbits observed to be slipping forward as a result of the curvature of space-time near the Sun. This same effect has been detected in pairs of stars that orbit each other.

A second test of general relativity was related to the motion of light through the curved space-time near the Sun. Because light has a limited speed, Newton's laws predict that the gravity of an object should slightly bend the paths of light beams passing nearby. The equations of general relativity indicated that light should have an extra deflection caused by traveling through curved space-time, just as a rolling golf ball is deflected by undulations in a putting green. Einstein predicted that starlight grazing the Sun's surface would be deflected by 1.75 arc seconds, twice the deflection that Newton's law of gravity would predict (**Figure 5-13**). Starlight passing near the Sun is normally lost in the Sun's glare, but during a total solar eclipse, stars beyond the Sun can be seen. As soon as Einstein published his prediction, astronomers rushed to observe such stars and test the curvature of space-time.

The first total solar eclipse following Einstein's publication of his theory of general relativity in 1916 occurred during June 1918. It was cloudy at some observing sites, and results from other sites were inconclusive. The next occurred in May 1919, only months after the end of World War I, and was visible from Africa and South America. British teams went to both Brazil and Príncipe, an island off the coast of Africa. Months before the eclipse, they photographed the part of the sky where the Sun would be located during the eclipse and measured the positions of the stars on the photographic plates. Then, during the eclipse, they photographed the same star field with the eclipsed Sun located in the middle. After measuring the plates, they found slight changes in the positions of the stars. Stars seen near the edge of the solar disk were observed to be shifted outward, away from the Sun, by about 1.8 arc seconds, close to the theory's prediction.

Because the angles are so small, this is a delicate observation, and it has been repeated at many total solar eclipses since 1919, with similar results (**Figure 5-14**). The most accurate results were obtained in 1973 when a combined University of Texas and Princeton team measured a deflection of 1.66 ± 0.18 arc seconds, which is in very good agreement with the prediction of Einstein's theory.

General relativity also predicts that a rapid change in a gravitational field should spread outward at the speed of light as a ripple in space-time labeled **gravitational radiation**. Decades ago, astronomers inferred the existence of gravitational radiation from details in the motions of some massive objects called *pulsars* that orbit extremely close together. In 2016, it was announced that the Laser Interferometer Gravitational Wave Observatory



▲ **Figure 5-14** (a) Schematic drawing of the deflection of starlight by the Sun's gravity. Dots show the true positions of the stars as photographed months before the eclipse. Lines point toward the positions of the stars during the eclipse. (b) Actual data from an eclipse in 1922. Random uncertainties of observation cause some scatter in the data, but on average the stars appear to move away from the Sun by 1.77 arc seconds at the edge of the Sun's disk. The deflection of stars is magnified by a factor of 1000 in both (a) and (b).

(LIGO) had made the first direct detection of gravitational radiation, providing one more confirmation of general relativity.

The general theory of relativity is critically important in modern astronomy. Einstein revolutionized modern physics by providing an explanation of gravity based on the geometry of curved space-time. Galileo's inertia and Newton's mutual gravitation are shown to be not just descriptive rules but fundamental properties of space and time.

Practicing **SCIENCE**

What does the equivalence principle tell you? Einstein began his work, as scientists sometimes do, by thinking carefully about common things such as what you feel when you are moving uniformly versus accelerating. This led him to deep insights, one of which is known as the *equivalence principle*.

The equivalence principle states that there is no observation you can make inside a closed spaceship to distinguish between uniform acceleration and gravitation. Of course, you could open a window and look outside, but then you would no longer be in a closed spaceship. As long as you make no outside observations, you can't tell whether your spaceship is firing its rockets and accelerating through space or resting on the surface of a planet where gravity gives you weight.

Einstein took the equivalence principle to mean that gravity and acceleration through space-time are somehow related. The general theory of relativity gives that relationship a mathematical form and shows that gravity is really a distortion in space-time that physicists refer to as *curvature*. Consequently, it is said that "mass tells space-time how to curve, and curved space-time tells mass how to move." The equivalence principle led Einstein to an explanation for gravity.

WHAT ARE WE?

Falling

Everything in the Universe is falling. The Moon is falling around Earth. Earth is falling along its orbit around the Sun, and the Sun and every other star in our Galaxy are falling along their orbits around the galactic center. Stars in other galaxies are falling around the center of those galaxies, and every galaxy in the Universe is falling as it feels the gravitational tugs of every bit of matter that exists.

Newton's explanation of gravity as a force between two unconnected masses was action at a distance, and it offended many of the scientists of his time. They thought Newton's gravity seemed like magic. In the 20th century, Einstein explained that gravity is a curvature of space-time and that every mass accelerates according to the curvature it feels around it. That's not action at a distance, and it gives new insight into how the Universe works.

The mass of every atom in the Universe contributes to the curvature, creating a universe filled with three-dimensional hills and valleys of curved space-time. You and Earth, the Sun, our Galaxy, and every other object in the Universe are falling through space guided by the curvature of space-time.

Study and Review

SUMMARY

5-1 Galileo's and Newton's Two New Sciences

- Aristotle argued that **natural motion** occurred when a displaced object returned to its proper place. **Violent motion** was motion other than natural motion and had to be sustained by a force. In fact, Galileo found that a falling object is accelerated; that is, it falls faster and faster with each passing second. The rate at which a falling object changes its speed, termed the **acceleration of gravity**, is 9.8 m/s^2 (32 ft/s^2) at Earth's surface and does not depend on the weight of the object. Resistance of matter to changes in motion is called **inertia**.
- Newton's first law of motion is that a body continues at rest or in uniform motion in a straight line unless it is acted on by some force. Newton's second law says that an **acceleration**, which is defined as a change in velocity, must be caused by a force and vice versa. A **velocity** is a speed with a specific direction, so a change in speed *or* direction is an acceleration. **Mass** is the amount of matter in a body. **Momentum** is a measure of a body's amount of motion, which is a combination of its velocity and mass. Newton's third law says that forces occur in pairs acting in opposite directions.
- From his mathematical analysis of the Earth–Moon system, Newton was able to show that the force of gravity between two masses is proportional to the product of their masses and depends on distance with an **inverse square law**. That is, the force of gravity is inversely proportional to the square of the distance between the two masses. To explain how gravity can act at a distance, scientists now describe it as a **field**, which means that there is a strength and direction of gravitational force associated with every point in space.

5-2 Orbital Motion and Tides

- Objects in orbit around Earth are falling (being accelerated) toward Earth's center. If there is no friction, the object will fall around its orbit forever. Two objects in orbit around each other actually orbit their common **center of mass**.
- An object in a **closed orbit** follows an elliptical path. To follow a circular orbit, an object must orbit with **circular velocity**:

$$V_c = \sqrt{\frac{GM}{r}} \qquad \text{(Eq. 5-1a)}$$

  At a certain distance from Earth, a **geosynchronous satellite** can stay above a spot on Earth's equator as Earth rotates.
- If a body's velocity equals or exceeds the **escape velocity, V_e**, it will follow a parabola or hyperbola. Escape velocity is defined as:

$$V_e = \sqrt{\frac{2GM}{r}} \qquad \text{(Eq. 5-1b)}$$

  These orbits are termed **open orbits** because the object never returns to its starting place.
- Newton's laws of motion and of gravity explain Kepler's three laws of planetary motion. The planets follow elliptical orbits because gravity obeys the inverse square law. The planets move faster when closer to the Sun and slower when farther away because they conserve **angular momentum**. A planet's orbital period squared is proportional to its orbital radius cubed because the moving planet conserves energy:

$$P^2 = \left(\frac{4\pi^2}{GM}\right) r^3 \qquad \text{(Eq. 5-2)}$$

- **Energy** refers to the ability to produce a change. Kinetic energy is an object's energy of motion, and potential energy is the energy an object has because of its position. The unit of energy is the **joule (J)**.
- Tides are caused by differences in the force of gravity acting on different parts of a body. On Earth, there are two tidal bulges on Earth caused by the Moon's gravity. Tides produced by the Moon combine with tides produced by the Sun to cause extreme tides, called **spring tides**, at new and full moons. The Moon and Sun work against each other to produce the smallest tides, called **neap tides**, at quarter moons. Friction from tides can slow the rotation of a rotating object, and the gravitational pull of tidal bulges can make orbits change slowly.

5-3 Einstein and Relativity

- Einstein's first postulate of **special theory of relativity** is that observers cannot detect their uniform motion through space by internal tests, only by observation of outside objects. This leads to the second relativity postulate: The speed of light measured in a vacuum is a constant for all observers. A consequence of special relativity is that mass and energy are related. That relationship is expressed by the famous equation,

$$E = m_0 c^2 \qquad \text{(Eq. 5-3)}$$

- The **general theory of relativity** says that a gravitational field is a curvature of space-time caused by the presence of a mass. For example, Earth's mass curves space-time, and the mass of your body responds to that curvature by accelerating toward Earth's center.
- General relativity's prediction of the curvature of space-time was confirmed by observations of the slow advance in perihelion (precession) of the orbit of Mercury and by the deflection of starlight passing near the Sun observed during a 1919 total solar eclipse. Further observations confirming Einstein's general theory of relativity, including **gravitational radiation**, continue to be made up to the present day.

REVIEW QUESTIONS

1. According to Aristotle, if earth and water were displaced then they would return naturally to their proper place. Today, what do we call this Aristotelean natural motion?
2. Today, what do we call the Aristotelean violent motion?
3. Which of Kepler's or Newton's laws best describes Aristotelean violent motions?
4. Why would Aristotle's explanation of gravity not work if Earth is not the center of the Universe?
5. According to the principles of Aristotle, what part of the motion of a baseball pitched across home plate is natural motion? What part is violent motion?

6. If you drop a feather and a steel hammer at the same moment, they should hit the ground at the same instant. Why doesn't this work on Earth, and why does it work on the Moon? Will it work on Phobos, a moon of Mars?
7. What is the difference between mass and weight?
8. When a person says he gained weight, does he mean that he gained in mass, gravity, or both mass and gravity?
9. An astronaut working in space near the International Space Station says she feels weightless. What does she mean? Does the astronaut not have weight?
10. What is the difference between speed and velocity?
11. A car is on a circular off ramp of an interstate and is traveling at exactly 25 mph around the curve. Does the car have velocity? Does the car have acceleration? Is the car decelerating?
12. How many accelerators does a car have? What are they?
13. You put your astronomy textbook and your No. 2 pencil on a ceramic tile floor, and you blow on each. Which has more inertia—the pencil, the textbook, or neither? Why? Which has more momentum? Why?
14. An astronaut is in space with a baseball and a bowling ball. The astronaut gives both objects an equal push in the same direction. Does the baseball have the same inertia as the bowling ball? Why? Does the baseball have the same acceleration as the bowling ball from the push? Why? If both balls are traveling at the same speed, does the baseball have the same momentum as the bowling ball?
15. You are at a red light in your car. The red light turns green, and you take your foot off the brake and step on the gas pedal, without turning the steering wheel. The speedometer changes from 0 to 20 mph in 3 seconds. In those 3 seconds, did the car change velocity? Is that an acceleration? Now, at a constant speed of 20 mph, you make a right turn. Is that an acceleration?
16. You weigh 100 pounds, your friend weighs 200 pounds, and you are in an arm wrestling contest with each other. Neither person is winning, but each of you is struggling to push the other's forearm over to the tabletop. Which of Newton's laws applies to this scenario and why? Now in one swift motion you plant your friend's forearm on the table, winning the contest. Which of Newton's laws applies to this motion and why?
17. Why did Newton conclude that some force had to pull the Moon toward Earth?
18. Why did Newton conclude that gravity has to be mutual and universal?
19. You have the same mass as a person sitting next to you. Are you gravitationally attracted to them? If so, why don't you instantly zoom over and stick to them? Is the attraction mutual? If not, why not?
20. You are sitting next to a person who has twice as much weight. Are you gravitationally attracted to them? If so, is it twice as much as their attraction to you? If not, why not?
21. You are sitting next to a person who has twice as much weight. You get up and move one seat over, doubling your distance from them. Did the gravitational force between you increase, decrease, or stay the same?
22. You are sitting next to a person who has twice as much weight as you do. A friend comes by and gives you a marshmallow, and you eat it. Did the gravitational force between you and your friend increase, decrease, or stay the same? Why?
23. How does the concept of a field explain action at a distance? Name another kind of field also associated with action at a distance.
24. Why can't a spacecraft go "beyond Earth's gravity"?
25. Where is the center of mass of your body?
26. Balance a pencil lengthwise on the side of your finger. Where is the center of mass? Balance a pencil widthwise (for example, on the eraser side) on your finger. Where is the center of mass? Is the center of mass a plane, sphere, circle, point, or a line?
27. What is the center of mass of two bodies? Where is the center of mass of the Earth–Moon system?
28. Why can't you leave Earth's gravitational field when jumping vertically?
29. According to Kepler's first law, planets move in elliptical orbits. Why is that considered accelerated motion? According to Newton, what is the force causing that acceleration?
30. How do planets orbiting the Sun and skaters doing a spin both conserve angular momentum?
31. If a planet were to slowly migrate inward toward the Sun while remaining in a stable orbit, would its orbital speed increase, decrease, or stay the same? Would its angular momentum change? Which of Kepler's laws or Newton's version of Kepler's laws does this scenario describe?
32. If you hold this textbook out at shoulder height and let go, at the instant you let go, does the book have potential energy? Kinetic energy?
33. Today at the beach you see the highest of all high tides during the last month. You see the Moon in the daytime sky. What is the most likely Moon phase? Why?
34. Why is the period of an open orbit undefined?
35. In what conditions do Newton's laws of motion and gravity need to be modified?
36. How does the first postulate of special relativity imply the second postulate?
37. When you ride a fast elevator upward, you feel slightly heavier as the trip begins and slightly lighter as the trip ends. How is this phenomenon related to the equivalence principle?
38. From your knowledge of general relativity, would you expect radio waves from distant galaxies to be deflected as they pass near the Sun? Why or why not?
39. How is gravity related to acceleration? Are all accelerations the result of gravity?
40. Near a massive planet, is gravitational acceleration large or small? Is space strongly curved, or not? What about near a small marble?
41. **How Do We Know?** Why would science be impossible if some natural events happened without causes?
42. **How Do We Know?** Why is it important that a theory make testable predictions?

ACTIVE INQUIRY QUESTIONS

1. Give an example from everyday life to illustrate each of Newton's laws.
2. People who lived before Newton may not have believed in cause and effect as strongly as you do. How do you suppose that affected how they saw their daily lives?
3. If Newton modified Kepler's laws and Einstein modified Newton's laws, is it possible Einstein's laws (postulates) might be modified?

PROBLEMS

1. An astronomy textbook is to be dropped from a tall building on Earth. One second after dropped, what are the textbook's speed, velocity, and acceleration? After 2 seconds? After 3 seconds? The book hits the ground; what are the book's speed, velocity, and acceleration?

2. Compared to the strength of Earth's gravity at its surface $r = R_E$ where R_E is the radius of Earth, how much weaker is gravity at a distance of $r = 10\ R_E$? At $r = 20\ R_E$?
3. Compare the force of gravity on a 1 kg mass on the Moon's surface with the force that mass on Earth's surface. Which force is greater, why, and by how much?
4. A satellite is in orbit at a distance r from the center of Earth. If the orbit radius is halved so that the satellite is orbiting closer to Earth's surface, will the field strength increase, decrease, or stay the same? If it changes, by how much?
5. The International Space Station is in orbit around Earth at a distance r from the center of Earth. A recent addition increased the Station's mass by a factor of 3. Did Earth's gravitational force on the Station increase, decrease, or stay the same? If it changes, by how much?
6. If a small lead ball falls from a high tower on Earth, what will be its velocity after 2 seconds? After 4 seconds?
7. What is the circular velocity of an Earth satellite 1000 km above Earth's surface? (*Note:* Earth's average radius is 6370 km. *Hint*: Convert all quantities to m, kg, s.)
8. What is the circular velocity of an Earth satellite 36,000 km above Earth's surface? What is its orbital period? (*Note:* Earth's average radius is 6370 km. *Hint:* Convert all quantities to m, kg, s.)
9. What is the orbital speed at Earth's surface? Ignore atmospheric friction. (*Note:* Earth's average radius is 6370 km. *Hint:* Convert all quantities to m, kg, s.)
10. Describe the shape of the orbit followed by the slowest cannonball in Figure 3 of the Concept Art, pretending that the cannonball could pass freely through Earth. (Newton got this problem wrong the first time he tried to solve it.)
11. If you visited a spherical asteroid 30 km in radius with a mass of 4.0×10^{17} kg, what would be the circular velocity at its surface? A major league fastball travels about 90 mph. Could a good pitcher throw a baseball into orbit around the asteroid? (*Note:* 90 mph is 40 m/s.)
12. What is the orbital period of a satellite orbiting just above the surface of the asteroid in Problem 11?
13. What would be the escape velocity at the surface of the asteroid in Problem 11? Could a major league pitcher throw a baseball off the asteroid so that it never came back?
14. Pretend you are an asteroid. What is the escape velocity from you, assuming your mass is 60 kg and your radius is 0.8 meters? Estimate how fast a fly flies. Would a fly be able to reach your escape velocity?
15. A moon of Jupiter takes 1.8 days to orbit at a distance of 4.2×10^5 km from the center of the planet. What is the total mass of Jupiter plus this moon? (*Note:* One day is 86,400 seconds.)

SENSE OF PROPORTION

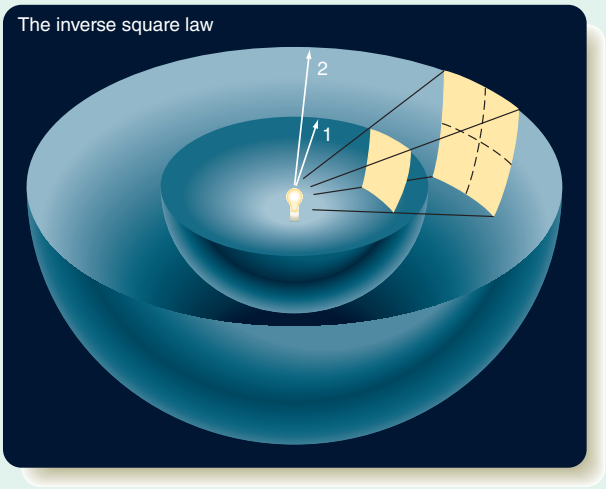
1. Arrange the following motions in order of increasing velocity: orbit of Earth around the Sun; an orbit around Earth near Earth's surface; an object just barely escaping Earth; orbit of the Moon around Earth; orbit of a geosynchronous satellite
2. Arrange the following motions in order of increasing acceleration: a dragster car; an object dropped off a cliff near the top of Mt. Everest; an object dropped by an astronaut standing on the Moon; an object dropped from the top of a New York City skyscraper

LEARNING TO LOOK

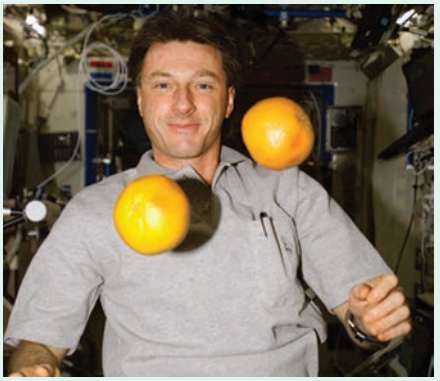
1. Why can the object shown here be bolted in place and used 24 hours a day without adjustment?

Larry Mulvehill/The Image Works

2. What is the flux at position 2 compared to position 1 in Figure 5-5? How does the distance from the center to position 2 compare with the distance to position 1?



3. Why is it a little bit misleading to say that this astronaut is weightless?


NASA

Light and Telescopes

6

ESO/B. Tafreshi (twanight.org)

▲ A portion of the ALMA millimeter/submillimeter radio telescope array, at work on a high plateau in the Chilean Andes. Each dish is 12 meters (40 ft) in diameter. The photograph's long exposure shows star trails caused by Earth's rotation. In the Southern Hemisphere, stars appear to circle around the south celestial pole that lies in the faint constellation of Octans (the Octant), not marked by any bright star.

In previous chapters of this book, you viewed the sky the way the first astronomers did, with the unaided eye. Then, in Chapter 4, you got a glimpse through Galileo's small telescope that revealed amazing things about the Moon, Jupiter, and Venus. Now you can consider the telescopes, instruments, and techniques of modern astronomers.

Telescopes gather and focus light, so you need to study what light is, and how it behaves, on your way to understanding how telescopes work. You will learn about telescopes that capture invisible types of light such as radio waves and X-rays. These enable astronomers to reach a more complete understanding of the Universe. This chapter will help you answer these five important questions:

- What is light?
- How do telescopes work?
- What are the powers and limitations of telescopes?
- What kind of instruments do astronomers use to record and analyze light gathered by telescopes?
- Why are some telescopes located in space?

Science is based on observations. Astronomers cannot visit distant stars and galaxies, so they must study them using telescopes. Many chapters of exploration remain, and every one of them will present information gained by astronomers using telescopes.

The strongest thing that's given
us to see with's
A telescope. Someone in every
town
Seems to me owes it to the
town to keep
one.

ROBERT FROST, "THE STAR-SPLITTER"
Reprinted by permission of Henry Holt & Company

LIGHT FROM THE SKY is a treasure that links you to the rest of the Universe. Astronomers strive to study light from the Sun, planets, moons, asteroids, comets, stars, nebulae, and galaxies, extracting information about their natures. Most celestial objects are very faint sources of light, so large telescopes are built to collect the greatest amount of light possible.

Some types of telescopes, for example radio telescopes like the ones featured on the previous page, gather light that is invisible to the human eye, but all telescopes work by the same basic principles. Some telescopes are used on Earth's surface, but others must go high in Earth's atmosphere, or even above the atmosphere into space, to work properly.

There is more to astronomy than amazing technology and brilliant scientific analysis. Astronomy helps us understand what we are. In the quotation that opens this chapter, the poet Robert Frost suggests that someone in every town should own a telescope to help us look upward and outward.

6-1 Radiation: Information from Space

Astronomers no longer spend their time mapping constellations or charting phases of the Moon. Modern astronomers analyze light using sophisticated instruments and techniques to investigate the temperatures, compositions, and motions of celestial objects and then make inferences about their internal processes and evolution. To understand how astronomers gain such detailed information about distant objects, you first need to learn about the nature of light.

6-1a Light as Waves and Particles

When you admire the colors of a rainbow, you are seeing an effect of light acting as a wave (Figure 6-1a). When you use a digital camera to take a picture of the same rainbow, the light acts like particles as it hits the camera's detectors (Figure 6-1b). Light has both wavelike and particle-like properties, and how it behaves depends partly on how you treat it.

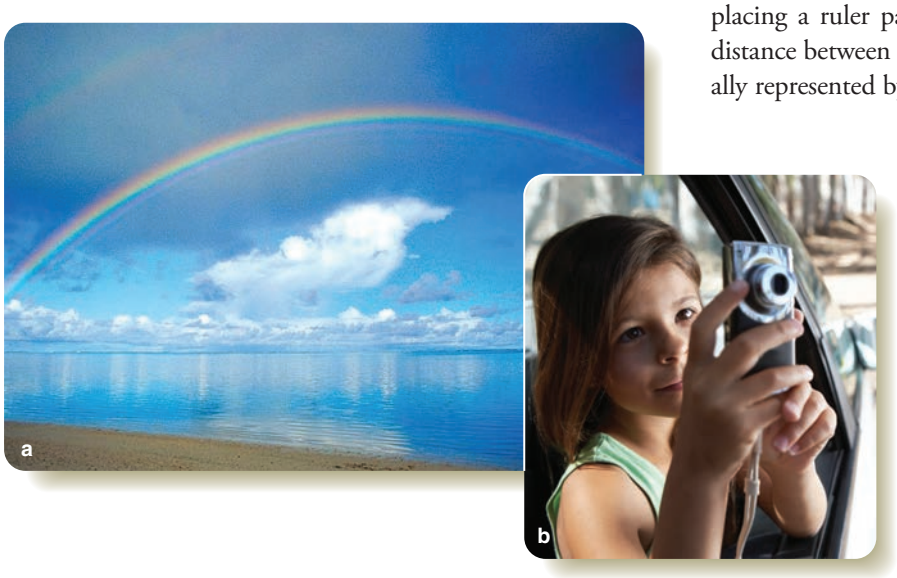
Light is referred to as **electromagnetic radiation** because it is made up of both electric and magnetic fields. (You encountered the concept of a field in the previous chapter in the context of gravitational fields.) The word *light* is commonly used to refer to electromagnetic radiation that humans can see, but visible light is only one among many types of electromagnetic radiation that include X-rays and radio waves.

Some people are wary of the word *radiation,* but that involves a **Common Misconception**. *Radiation* refers to anything that radiates away from a source. Dangerous high-energy particles emitted from radioactive atoms are also called radiation, and you have learned to be concerned when you hear that word. But light, like all electromagnetic radiation, spreads outward from its origin, so you can correctly refer to light as a form of radiation.

Electromagnetic radiation travels through space at a speed of 3.00×10^8 m/s (186,000 mi/s). This is commonly referred to as the speed of light, symbolized by the letter *c,* but it is in fact the speed of all types of electromagnetic radiation.

Electromagnetic radiation can act as a wave phenomenon, associated with a periodically repeating disturbance—a wave—that carries energy. You are familiar with waves in water: If you disturb a pool of water, waves spread across the surface. Imagine placing a ruler parallel to the travel direction of the wave. The distance between peaks of the wave is called the **wavelength**, usually represented by the Greek lowercase letter *lambda* (λ).

Wavelength is related to **frequency**, the number of waves that pass a stationary point in 1 second. Frequency is often represented by the Greek lowercase letter *nu* (ν). The



Panel a: © Gail Johnson/Shutterstock.com; b: Roy McMahon/Cardinal/Corbis

◀ **Figure 6-1** The wavelike properties of light produce a rainbow, whereas the particle-like properties are involved in the operation of a digital camera.

relationships among the wavelength, frequency, and speed of a wave can be expressed by:

KEY EQUATION 6-1 Wavelength–Frequency–Speed

$$\lambda\nu = c$$

Example: Let's say your favorite FM station is on the dial at 89.5, which means that radio station is producing waves with a frequency $\nu = 89.5$ megahertz. In other words, 89.5 million radio wave peaks from this station pass by you each second. What is the wavelength of those radio waves? ***Solution***: You already know that the radio waves are traveling at the speed of light $c = 3.00 \times 10^8$ m/s. Using Eq. 6-1, you can calculate that this particular radio station is radiating radio waves with a wavelength of $\lambda = 3.35$ m. Note that wavelength and frequency have an inverse relationship: The higher the frequency, the shorter the wavelength.

Sound is another example of a wave—in this case, a periodically repeating pressure disturbance that moves from source to ear. Sound requires a medium, meaning a substance such as air, water, or rock to travel through. In contrast, light is made up of electric and magnetic fields that do not require a medium and can travel through empty space. For example, on the Moon, where there is no air, there can be no sound, but there is plenty of light. This brings up a **Common Misconception** that radio waves are related to sound. Actually, radio waves are a type of light (electromagnetic radiation) that your radio receiver transforms into sound so you can listen. Radio communication works just fine between astronauts standing on the airless Moon; radio signals travel through the vacuum, and then the spacesuit radios convert the radio signals to sound that is heard in the air inside their helmets.

Although electromagnetic radiation can behave as a wave, it can also behave as a stream of particles. A particle of electromagnetic radiation is called a **photon**. You can think of a photon as a packet of waves. The amount of energy a photon carries is inversely proportional to its wavelength. The following simple formula describes that relationship:

$$E = \frac{hc}{\lambda}$$

Here h is Planck's constant (6.63×10^{-34} joule s), c is the speed of light in meters per second, and λ is the wavelength in meters.

This equation expresses the important point that there is a relationship between the energy E of a photon that is a particle property of light, and the wavelength λ that is a wave property. The inverse proportion means that as λ gets smaller E gets larger: Shorter-wavelength photons carry more energy, and longer-wavelength photons carry less energy. You can see that the relationship between wavelength and frequency means there must also be a simple relationship between photon energy and frequency. That is, short wavelength, high frequency, and large photon energy go together; long wavelength, low frequency, and small photon energy go together.

6-1b The Electromagnetic Spectrum

A **spectrum** is an array of electromagnetic radiation displayed in order of wavelength. You are most familiar with the spectrum of visible light that you see in rainbows. The colors of the rainbow differ in wavelength, with red having the longest wavelength and violet the shortest.

The average wavelength of visible light is about 0.0005 mm. This means that roughly 50 light waves would fit end to end across the thickness of a sheet of household plastic wrap. It is awkward to describe such short distances in millimeters, so scientists usually give the wavelength of light using **nanometer (nm)** units, equal to one-billionth of a meter (10^{-9} m). Another unit that astronomers commonly use is called the **angstrom (Å)**, named after the Swedish astronomer Anders Jonas Ångström. One angstrom is 10^{-10} m, that is, one-tenth of a nanometer. The wavelength of visible light ranges from about 400 to 700 nm (4000 to 7000 Å).

Just as you sense the wavelength of sound as pitch, you sense the wavelength of light as color. Light with wavelengths at the short-wavelength end of the visible spectrum (λ = about 400 nm) appears violet to your eyes, and light with wavelengths at the long-wavelength end (λ = about 700 nm) appears red.

Infrared (IR) radiation was discovered in the year 1800, the first known example of "invisible light" (**Figure 6-2**). Infrared

Print Collector/Getty Images

▲ **Figure 6-2** Depiction of Sir William Herschel discovering that sunlight contains radiation detectable by thermometers but not by human eyes. He named that invisible light "infrared," meaning "below red."

radiation has wavelengths ranging from 700 nm to about 1 mm (1 million nm). Your eyes do not detect infrared radiation, but your skin can sense some of it as heat. A heat lamp warms you by giving off infrared radiation. Infrared radiation has longer wavelengths than the light your eyes can see.

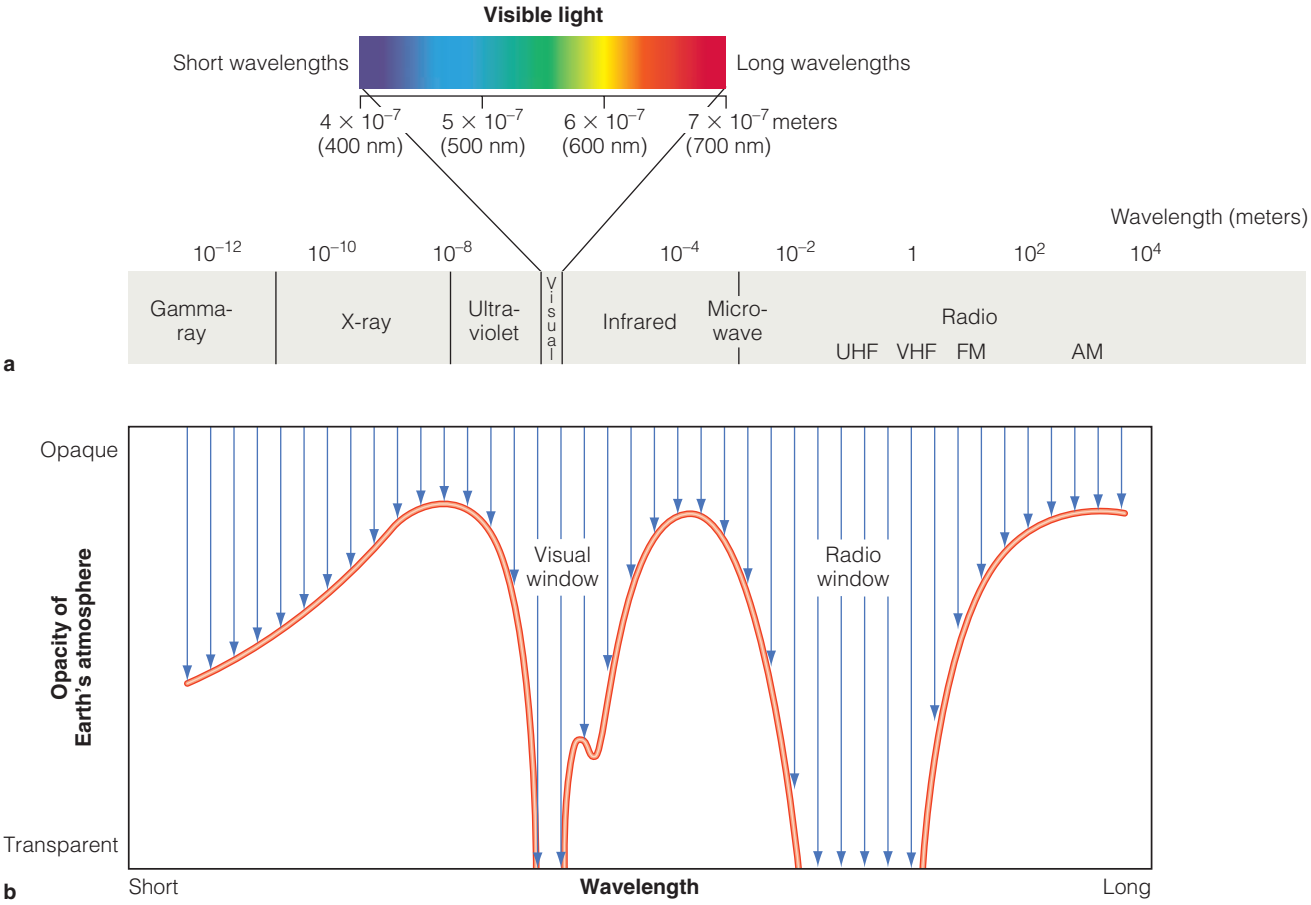
The visible spectrum is now known to be just a small part of the electromagnetic spectrum (Figure 6-3a). Beyond the infrared part of the electromagnetic spectrum lie **microwaves** and **radio waves**. Microwaves, used for cooking food in a microwave oven, as well as for radar and some long-distance telephone communications, have wavelengths from a few millimeters to a few centimeters. The radio waves used for FM, television, military, government, and cell phone radio transmissions have wavelengths of a few centimeters to a few meters, whereas AM and other types of radio transmissions have wavelengths of a few hundred meters to a few kilometers.

Now look at the other end of the electromagnetic spectrum in Figure 6-3a and notice that electromagnetic waves shorter than violet are called **ultraviolet (UV) radiation**. Electromagnetic waves that are even shorter are called **X-rays**, and the shortest are **gamma-rays**.

Recall the formula for the energy of a photon. Extremely short-wavelength, high-frequency photons, such as X-rays and gamma-rays, have high energies and can be dangerous. Even ultraviolet photons have enough energy to harm you. Small amounts of ultraviolet radiation produce a suntan, and larger doses cause sunburn and skin cancers. Contrast this to the lower-energy infrared photons. Individually they have too little energy to affect skin pigment, a fact that explains why you can't get a tan from a heat lamp. Only by concentrating many low-energy photons in a small area, as in a microwave oven, can you transfer significant amounts of energy.

The boundaries between these wavelength ranges are defined only by conventional usage, not by natural divisions. There is no real distinction between short-wavelength ultraviolet light and long-wavelength X-rays. Similarly, long-wavelength infrared radiation is indistinguishable from short-wavelength microwaves.

Astronomers collect and study electromagnetic radiation from space because it carries almost the only clues available about the nature of stars, planets, and other celestial objects. Earth's atmosphere is opaque to most electromagnetic radiation, as shown in the graph in Figure 6-3b. Gamma-rays and X-rays are absorbed high in Earth's atmosphere, and a layer of ozone (O_3) at altitudes of about 15 to 30 km (10 to 20 mi) absorbs most ultraviolet radiation. Water vapor in the lower atmosphere absorbs most long-wavelength infrared radiation and microwaves. Only visible light, some short-wavelength infrared radiation, and some radio waves reach Earth's surface through wavelength bands called **atmospheric windows**. Obviously, if you wish to study the Universe from Earth's surface, you have to "look through" one of those windows.



▲ **Figure 6-3** (a) The spectrum of visible light, extending from red to violet, is only part of the electromagnetic spectrum. (b) Most forms of light (electromagnetic radiation) are absorbed in Earth's atmosphere. Light can reach Earth's surface only through the visual and radio "windows."

Practicing SCIENCE

What would you see if your eyes were sensitive only to radio wavelengths? If you had radio vision, you would probably be able to see through walls because ordinary walls are transparent to most radio wavelengths. What you would see through the walls would be the many strong radio wave sources on Earth—radio and TV stations, cell phones, power lines, and even electric motors.

As you have learned, Earth's atmosphere is mostly transparent to radio waves. If, after looking around the surface of Earth, you looked up at the sky, you would see the Sun and Jupiter, which are both strong natural radio sources, but probably nothing else in the Solar System. You would also see numerous unfamilliar radio stars because few if any of the stars that are bright at visual wavelengths are also strong radio sources.

Now imagine a slightly different situation. ***Would you be in the dark if your eyes were sensitive only to X-ray wavelengths?***

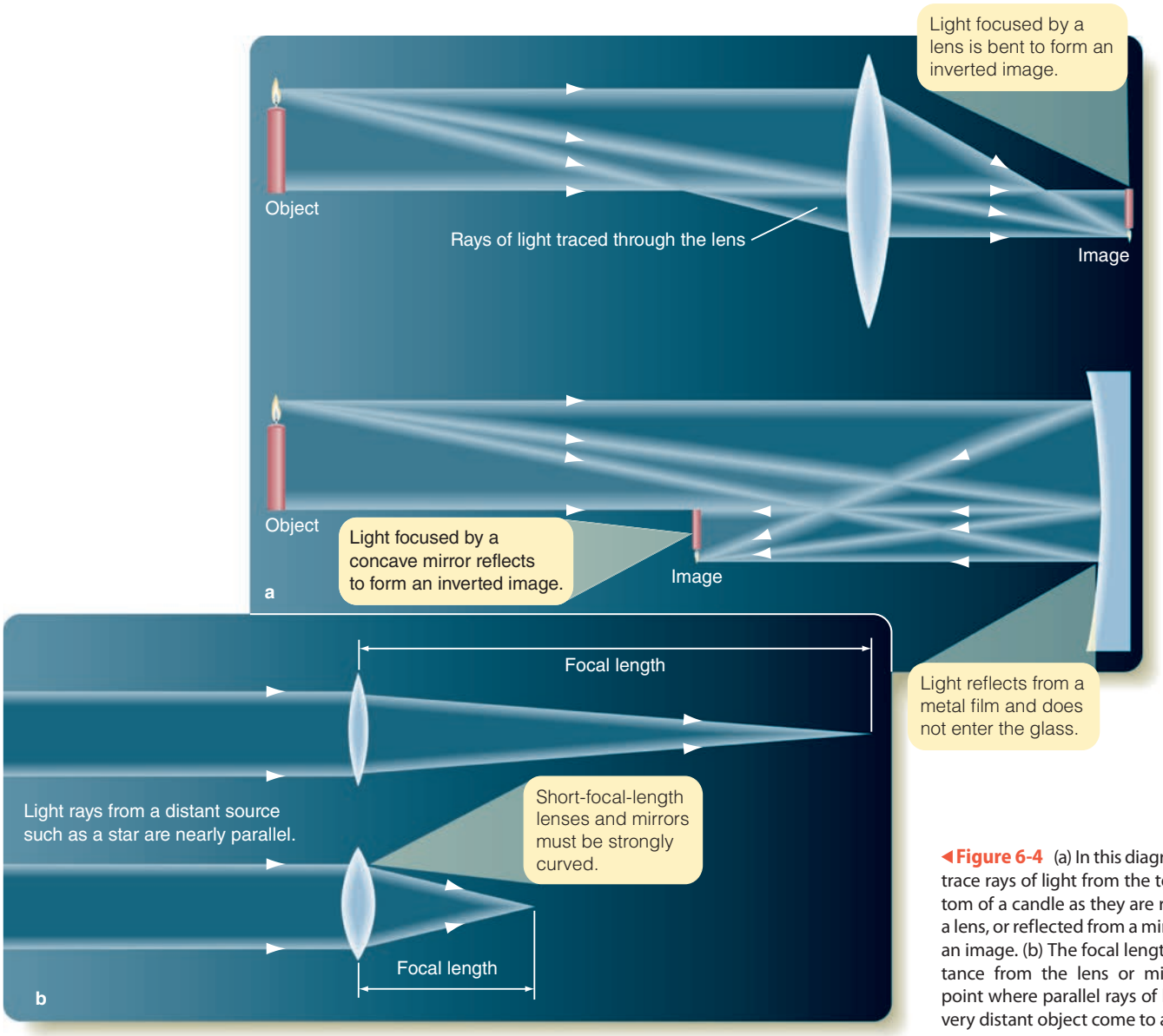
6-2 Telescopes

Astronomers build telescopes to collect light from distant, faint objects for analysis. That requires very large telescopes built by careful optical and mechanical engineering work. You can understand these ideas more completely by learning about the two types of telescopes and their relative advantages and disadvantages.

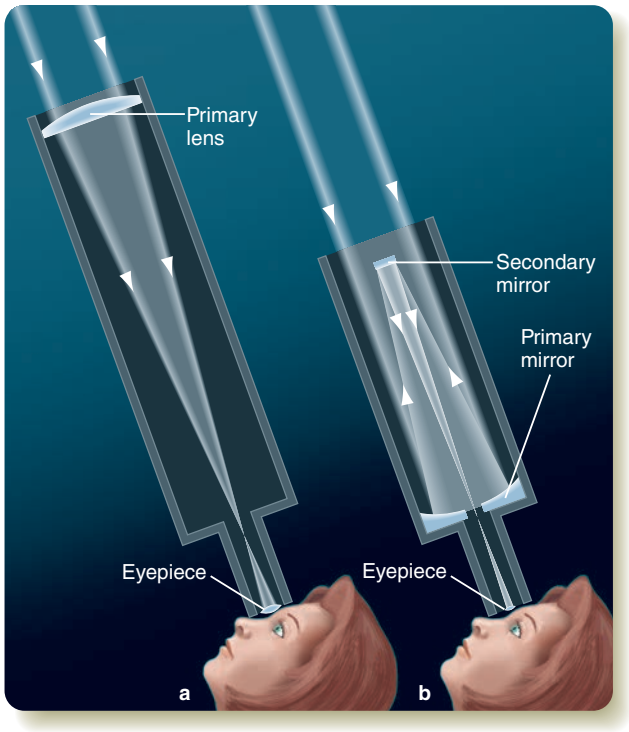
6-2a Refracting and Reflecting Telescopes

Light can be focused into an image in one of two ways (**Figure 6-4**). Either (1) a lens refracts ("bends") light passing through it, or (2) a mirror reflects ("bounces") light from its surface.

These two ways to manipulate light correspond to two astronomical telescope designs. **Refracting telescopes** use a lens to gather and focus light, whereas **reflecting telescopes** use a mirror (**Figure 6-5**). You learned in Chapter 4 that Galileo was the first person to systematically record observations of celestial objects using a telescope,


Light focused by a lens is bent to form an inverted image.
Object
Rays of light traced through the lens
Image
Object
Light focused by a concave mirror reflects to form an inverted image.
Image
a
Light reflects from a metal film and does not enter the glass.
Focal length
Light rays from a distant source such as a star are nearly parallel.
Short-focal-length lenses and mirrors must be strongly curved.
Focal length
b

◀ **Figure 6-4** (a) In this diagram you can trace rays of light from the top and bottom of a candle as they are refracted by a lens, or reflected from a mirror, to form an image. (b) The focal length is the distance from the lens or mirror to the point where parallel rays of light from a very distant object come to a focus.



▲ **Figure 6-5** (a) A refracting telescope uses a primary lens to focus starlight into an image that is magnified by another lens called an eyepiece. The primary lens has a long focal length, and the eyepiece has a short focal length. (b) A reflecting telescope uses a primary mirror to focus the light by reflection. In this particular reflector design, called a Cassegrain telescope, a small secondary mirror reflects the starlight back down through a hole in the middle of the primary mirror to the eyepiece lens.

beginning a little more than 400 years ago in 1610. Galileo's telescope was a refractor. In Chapter 5 you learned about the amazing range of Isaac Newton's scientific work; among his many accomplishments was the invention of the reflecting telescope.

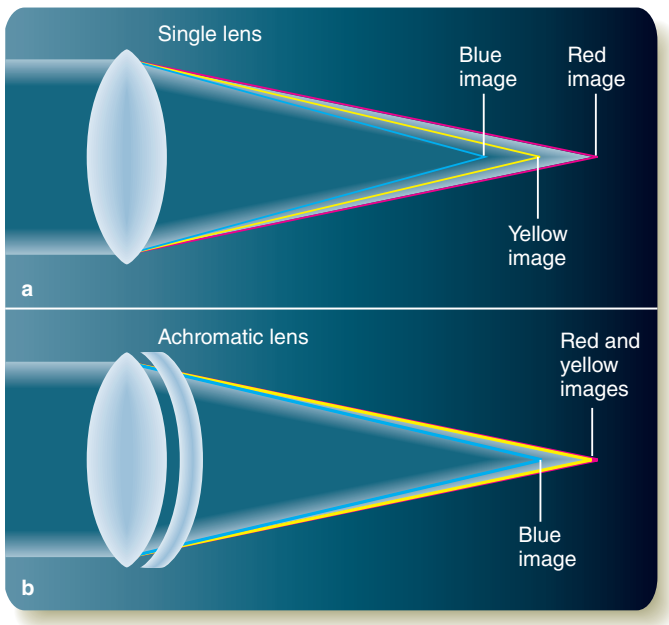
The main lens in a refracting telescope is called the **primary lens**, and the main mirror in a reflecting telescope is called the **primary mirror**. The distance from a lens or mirror to the image it forms of a distant light source such as a star is called the **focal length**. Both refracting and reflecting telescopes form an image that is small, inverted, and difficult to observe directly, so a lens called the **eyepiece** normally is used to magnify the image and make it convenient to view.

Manufacturing a lens or mirror to the proper shape and necessary smoothness is a delicate, time-consuming, and expensive process. Short focal-length lenses and mirrors must be made with more curvature than ones with long focal lengths. The surfaces of lenses and mirrors then must be polished to eliminate irregularities larger than the wavelengths of light. Creating the optics for a large telescope can take months or years; involve huge, precision machinery; and employ several expert optical engineers and scientists.

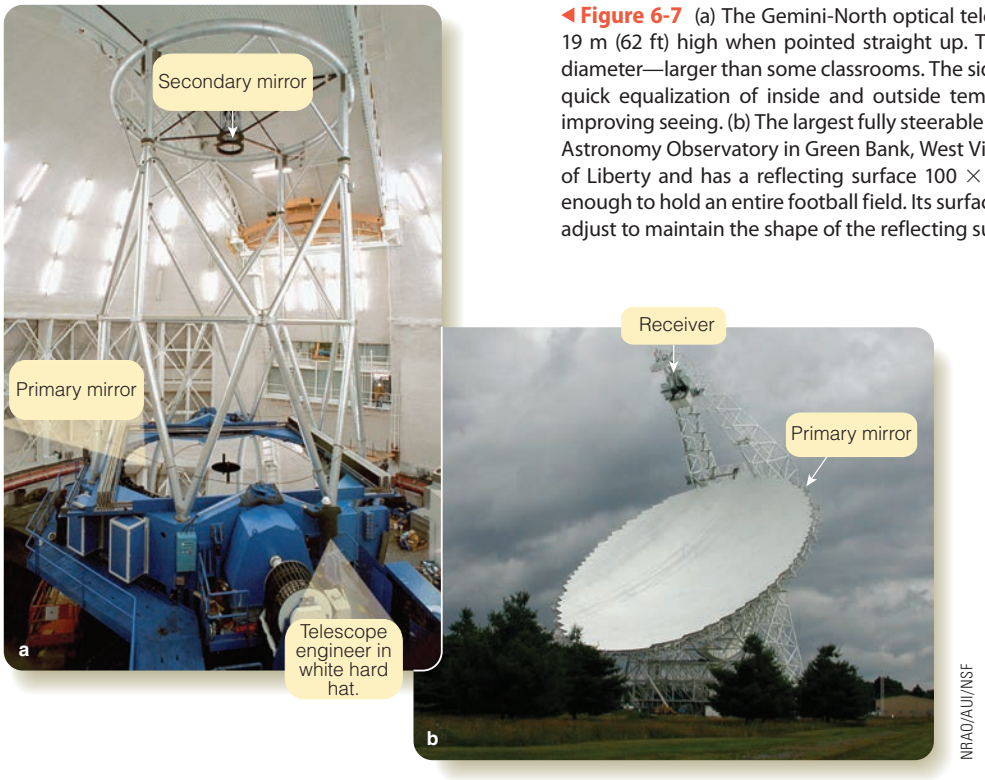
Refracting telescopes suffer from a serious optical distortion that limits their use. When light is refracted through glass, shorter-wavelength light bends more than longer wavelengths; so, for example, blue light comes to a focus closer to the lens than does red light (**Figure 6-6a**). That means if you focus the eyepiece on the blue image, the other colors are out of focus, and you see a colored blur around the image. If you focus instead on the red image, all the colors except red are blurred, and so on. This color separation is called **chromatic aberration**. Telescope designers can grind a telescope lens with two components made of different kinds of glass and thereby bring two different wavelengths to the same focus (**Figure 6-6b**). That improves the image, but these so-called **achromatic lenses** are not totally free of chromatic aberration. Even though two colors have been brought together, the others are still out of focus.

A refracting telescope's primary lens is much more difficult to manufacture than a mirror of the same size. The interior of the glass must be pure and flawless because the light passes through it. Also, if the lens is achromatic, it must be made of two different kinds of glass requiring four precisely ground surfaces. The largest refracting telescope in the world was completed in 1897 at Yerkes Observatory in Wisconsin. Its achromatic primary lens has a diameter of 1 m (40 in.) and weighs half a ton. Refracting telescopes larger than that would be prohibitively expensive.

The primary mirrors of reflecting telescopes are much less expensive than lenses because the light reflects off the front surface of the mirror. This means that only the front surface needs to be made with a precise shape and that surface is coated with a highly reflective surface of aluminum or silver. Consequently, the glass of the mirror does not need to be transparent, and the mirror can be supported across its back surface to reduce sagging caused by its own weight. Most important, reflecting telescopes do not suffer from chromatic aberration because the light does not pass through the glass, so reflection does not depend on wavelength. For these



▲ **Figure 6-6** (a) An ordinary lens suffers from chromatic aberration because short wavelengths bend more than long wavelengths. (b) An achromatic lens, with two components made of two different kinds of glass, can bring any two colors to the same focus, but other colors remain slightly out of focus.



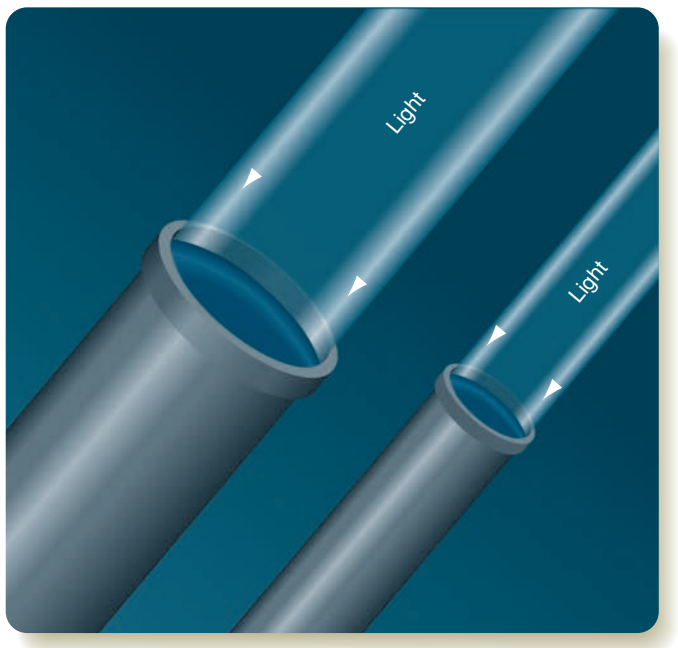
Gemini Telescope/AURA/NSF

NRAO/AUI/NSF

◀ **Figure 6-7** (a) The Gemini-North optical telescope on Mauna Kea in Hawai'i stands more than 19 m (62 ft) high when pointed straight up. The primary mirror (at bottom) is 8.1 m (26.5 ft) in diameter—larger than some classrooms. The sides of the telescope dome can be opened, allowing quick equalization of inside and outside temperatures at sunset, reducing air turbulence and improving seeing. (b) The largest fully steerable radio telescope in the world is at the National Radio Astronomy Observatory in Green Bank, West Virginia. The telescope stands higher than the Statue of Liberty and has a reflecting surface 100 × 110 m (330 × 360 ft) in diameter, more than big enough to hold an entire football field. Its surface consists of 2004 computer-controlled panels that adjust to maintain the shape of the reflecting surface.

reasons, all large astronomical telescopes built since the start of the 20th century have been reflecting telescopes.

Telescopes intended for the study of visible light are called **optical telescopes** (**Figure 6-7a**). As you learned previously, radio waves as well as visible light from celestial objects can penetrate Earth's atmosphere and reach the ground. Astronomers gather radio waves using **radio telescopes** such as the one in **Figure 6-7b** that resemble giant TV satellite dishes. It is technically extremely difficult to make a lens that can focus radio waves, so all radio telescopes, including small ones, are reflecting telescopes; the dish is the primary mirror.



▲ **Figure 6-8** Gathering light is like catching rain in a bucket. A large-diameter telescope gathers more light and produces a brighter image than a smaller telescope of the same focal length.

6-2b The Powers and Limitations of Telescopes

A telescope's capabilities are described in three mathematical relationships that are called the three powers of a telescope. The two most important of these powers depend on the diameter of the telescope.

Light-Gathering Power Nearly all of the interesting objects in the sky are faint sources of light, so astronomers need telescopes that can collect large amounts of light to be able to study those objects. **Light-gathering power** refers to the ability of a telescope to collect light. Catching light in a telescope is like catching rain in a bucket—the bigger the bucket, the more rain it can catch (**Figure 6-8**).

Light-gathering power is proportional to the *area* of the telescope primary lens or mirror; a lens or mirror with a large area gathers a large amount of light. The area of a circular lens or mirror written in terms of its diameter D is $\pi D^2/4$. To compare the relative light-gathering powers (LGP) of two telescopes A and B, you can calculate the ratio of the areas of their primaries, which equals the ratio of the primaries' diameters squared:

$$\frac{(LGP_A)}{(LGP_B)} = \left(\frac{D_A}{D_B}\right)^2$$

Example 1: Suppose you compare a 4-cm telescope with a 24-cm telescope. How much more light will the larger telescope gather? ***Solution:***

$$\frac{LGP_{24}}{LGP_4} = \left(\frac{24}{4}\right)^2 = 6^2 = 36 \text{ times more light}$$

Example 2: Your dark-adapted eye acts like a telescope with a diameter of about 0.8 cm, the maximum diameter of the pupil. How much more light can you gather if you use a 24-cm telescope, relative to your unaided eye? ***Solution:***

$$\frac{LGP_{24}}{LGP_{eye}} = \left(\frac{24}{0.8}\right)^2 = 30^2 = 900 \text{ times more light}$$

Because the diameter ratio is squared, even a small increase in diameter produces a relatively large increase in light-gathering power and allows astronomers to study significantly fainter objects. This principle holds not just at visual wavelengths but also for telescopes collecting any kind of radiation.

Resolving Power The second power of a telescope, called **resolving power**, refers to the ability of the telescope to reveal fine detail. One consequence of the wavelike nature of light is that there is an unavoidable blurring called **diffraction fringes** around every point of light in an image, and you cannot see any detail smaller than the fringes (**Figure 6-9**).

Astronomers can't eliminate diffraction fringes, but the size of the diffraction fringes is inversely proportional to the diameter of the telescope. This means that the larger the telescope, the better its resolving power. However, the size of diffraction fringes is also proportional to the wavelength of light being focused. In other words, an infrared or radio telescope has less resolving power than an optical telescope of the same size.

You can imagine testing the resolving power of a telescope by measuring the angular distance between two stars that are just barely distinguishable as separate objects (Figure 6-9b). The resolving power α in arc seconds of a telescope with primary diameter D that is collecting light of wavelength λ equals:

KEY EQUATION 6-2 Telescope Resolution

$$\alpha \text{ (arc seconds)} = 2.06 \times 10^5 \left(\frac{\lambda}{D}\right)$$

To use the formula correctly, the units of D and λ need to be the same, for example, meters and meters, or centimeters and centimeters. The multiplication factor of 2.06×10^5 is the conversion between radians and arc seconds that you first saw in the small-angle formula (Eq. 3-1). If the wavelength of light being studied is assumed to be 550 nm, in the middle of the visual band, then the preceding formula simplifies to:

$$\alpha \text{ (arc seconds)} = \frac{0.113}{D}$$

Example: What is the resolving power of a 10.0-cm (= 0.100 m, or about 4 in.) diameter telescope observing at visual wavelengths? ***Solution:***

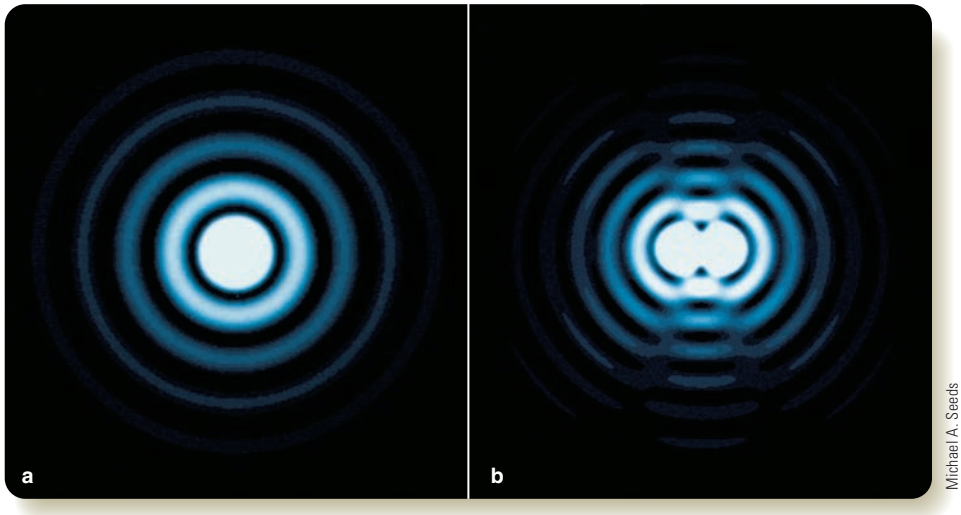
$$\alpha = 2.06 \times 10^5 \left(\frac{550 \times 10^{-9}}{0.100}\right) = 1.13 \text{ arc seconds}$$

Or, equivalently,

$$\alpha = \frac{0.113}{0.100} = 1.13 \text{ arc seconds}$$

In other words, using a 10-cm (about 4-in.)-diameter telescope, you should be able to distinguish as separate points of light any pair of stars farther apart than about 1.1 arc seconds if the lenses are of good quality and if the atmosphere is stable. Stars closer together than that will be blurred together into a single image by the diffraction fringes.

Aside from diffraction, two other factors—optical quality and atmospheric conditions—limit resolving power. A telescope must have high-quality optics to achieve its full potential resolving power. Even a large telescope reveals little detail if its optical surfaces are marred by imperfections. Also, when you look through a telescope, you are looking up through miles of turbulent air in Earth's atmosphere, which inevitably makes images wiggle and blur to some extent. Astronomers use the term **seeing** to refer to the amount of image wiggling and blurring as a result of atmospheric conditions. A related phenomenon is the twinkling of stars. Star twinkles are caused by turbulence in Earth's atmosphere, and a star near the horizon, where you look through more air, will twinkle and blur more than a star overhead.



Michael A. Seeds

Michael A. Seeds

◀ **Figure 6-9** (a) Stars are so far away that their images are points, but the wavelike characteristic of light causes each star image to be surrounded with diffraction fringes, much magnified in this computer model. (b) Two stars close to each other have overlapping diffraction fringes and become impossible to detect separately.

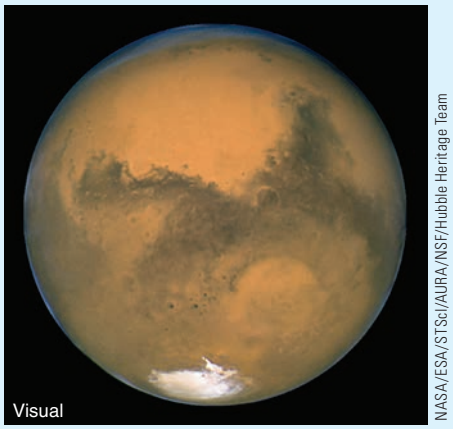
Resolution and Precision

What limits the precision of an observation?

As an example, think about observations that are in the form of images. All images have limited resolution. You can see on your computer screen that images there are made up of picture elements, pixels. If your screen has low resolution, it has large pixels, and you can't see much detail. In an astronomical image, the practical size of a pixel is set by the resolution limit, a combination of atmospheric seeing, telescope optical quality, and telescope diffraction. You can't see details smaller than the resolution limit. This limitation on the level of detail viewable in an image is one example of the limited precision, and therefore unavoidable uncertainty, of all scientific measurements.

Now imagine a zoologist trying to measure the length of a live snake by holding it along a meter stick. Meter sticks are usually not marked with resolution smaller than millimeters. Also, the wriggling snake is hard to hold, so it is difficult to measure accurately. Both factors—the meter stick's resolution and the snake's wriggling—together limit the precision of the measurement. If the zoologist said the snake is 432.8932 mm long, you might wonder if that is really true. The best resolution possible in that zoologist's situation does not justify the precision implied by all those digits. Images made with even the largest and best telescopes do not show surface details on stars because of limits on precision (resolution) set by diffraction and atmospheric seeing (a stellar equivalent of the snake wriggling).

If you are a scientist, one question you must ask yourself routinely is: How precise are the measurements you and other investigators have made? Precision of measurements is limited by the resolution of the measurement technique such as the size of the pixels in a photograph or the finest markings on a meter stick as much as by variability in what is being observed such as the snake wriggling or atmospheric turbulence. And, because precision is always limited, uncertainty is always present.

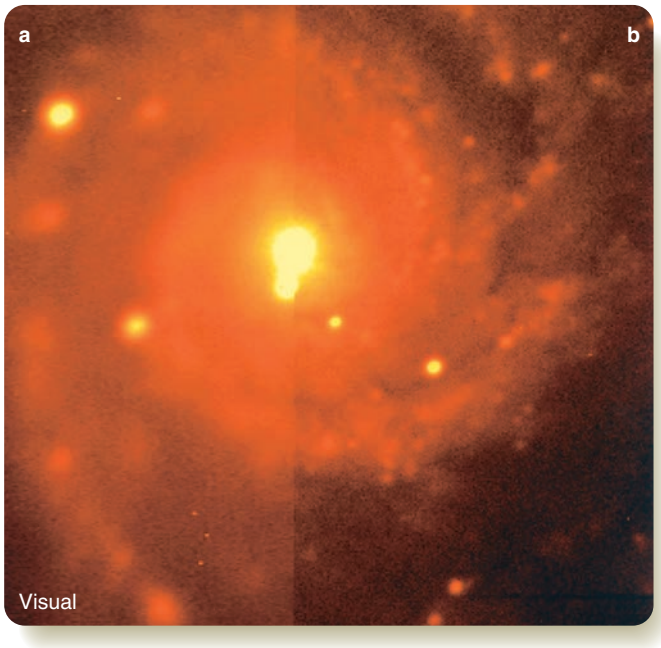


NASA/ESA/STScI/AURA/NSF/Hubble Heritage Team

A high-resolution visual-wavelength image of Mars made by the *Hubble Space Telescope* reveals details such as mountains, craters, and the south polar cap.

On a night when the atmosphere is unsteady, images are badly blurred, and astronomers say that "the seeing is bad" (**Figure 6-10a**). Generally, even under relatively good seeing conditions, the detail visible through a large telescope is limited not by its diffraction fringes but by the turbulence of the air through which the telescope must look. An optical telescope performs better on a high mountaintop where the air is thin and steady. But even in that situation, Earth's atmosphere spreads star images at visual wavelengths into blobs about 0.5 to 1.0 arc second in diameter. Radio telescopes are also affected by atmospheric seeing, but less than optical telescopes, so they do not benefit much in this respect by being located on mountains. You will learn later in this chapter about special techniques that improve seeing from ground-based telescopes and also about telescopes that orbit above Earth's atmosphere and are not limited by seeing.

Seeing and diffraction both limit the precision of any measurement that can be made using that image, and that limits the amount of information in the image. All measurements have some built-in uncertainty (**How Do We Know? 6-1**), and scientists must learn to work within those limitations. Have you ever tried to magnify a photograph to distinguish some detail? Photographs are composed of tiny picture elements, pixels, and no detail smaller than a single pixel will be visible no matter how much you magnify the photo. In an astronomical image, the resolution is limited by seeing, or diffraction, or both. You can't see any detail in the image that is smaller than the telescope's resolution. That's why stars look like fuzzy points of light no matter how big the telescope.



Courtesy William Keel

▲ **Figure 6-10** (a) The left half of this photograph of a galaxy is from an image recorded on a night of poor seeing. Small details are blurred. (b) The right half of the photo is from an image recorded on a night when Earth's atmosphere above the telescope was steady and the seeing was better. Much more detail is visible under good seeing conditions.

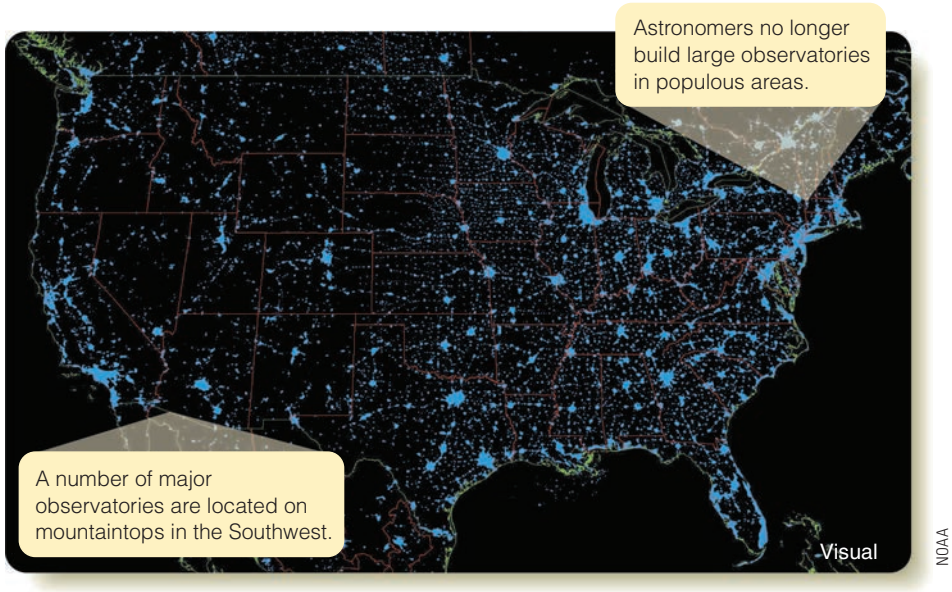
Magnifying Power The idea that the purpose of an astronomical telescope is to magnify images is a **Common Misconception**. In fact, the **magnifying power** of a telescope—its ability to make images bigger—is the least important of the three powers. Because the amount of detail that a telescope can discern is limited generally either by its resolving power or the seeing conditions, very high magnification does not necessarily show more detail. The magnifying power of a telescope equals the focal length of the primary mirror or lens divided by the focal length of the eyepiece.

$$M = \left(\frac{F_p}{F_e}\right)$$

Example: What is the magnification of a telescope with a primary mirror focal length of 80 cm if it is used with an eyepiece with a focal length of 0.50 cm? ***Solution:*** The magnification is 80 divided by 0.50, or 160 power.

Radio telescopes, of course, don't have eyepieces, but they do have instruments that examine the radio waves focused by the telescope, and each such instrument would, in effect, have its own magnifying power.

As was mentioned previously, the two most important powers of the telescope—light-gathering power and resolving power—depend on the diameter of the telescope that is essentially impossible to change. In contrast, you can change the magnification of a telescope simply by changing the eyepiece. This explains why astronomers describe telescopes by diameter and not by magnification. Astronomers will refer to a telescope as a 4-meter telescope or a 10-meter telescope, but they would never identify a research telescope as being, say, a 1000-power telescope.


Astronomers no longer build large observatories in populous areas.
A number of major observatories are located on mountaintops in the Southwest.
Visual
NOAA

▲ **Figure 6-11** This satellite view of the continental United States at night shows the light pollution and energy waste resulting from outdoor lighting. Observatories are best located far from large cities.

6-3 Observatories on Earth: Optical and Radio

The quest for light-gathering power and good resolution explains why nearly all the world's major observatories are located far from big cities and, especially in the case of optical telescopes, usually on top of mountains. Astronomers avoid cities because **light pollution**, which is the brightening of the night sky by light scattered from artificial outdoor lighting, can make it impossible to see faint objects (**Figure 6-11**). In fact, many residents of cities are unfamiliar with the beauty of the night sky because they can see only the brightest stars. Even far from cities, the Moon, nature's own light pollution, is sometimes so bright it drowns out fainter objects, and astronomers are unable to perform certain types of observations during nights near full moon. On such nights, faint objects cannot be detected even with large telescopes at good locations.

Radio astronomers face a problem of radio interference comparable to visible light pollution. Weak radio waves from the cosmos are easily drowned out by human-made radio noise—everything from automobiles with faulty spark plugs to poorly designed communication systems. A few narrow radio bands are reserved for astronomy research, but even those are often contaminated by stray signals. To avoid that noise and have the radio equivalent of a dark sky, astronomers locate radio telescopes as far from civilization as possible. Hidden in mountain valleys or in remote deserts, they are able to study the Universe protected from humanity's radio output.

As you have already learned, astronomers prefer to put optical telescopes on high mountains for several reasons. To find sites with the best seeing, astronomers carefully select mountains where the airflow is measured to be smooth and not turbulent. Also, the air at high altitude is thin, dry, and more transparent, which is important not only for optical telescopes but also for other types of telescopes. Building an observatory on top of a remote high mountain is difficult and expensive, as you can imagine from **Figure 6-12**, but the dark sky, good seeing, and transparent atmosphere make it worth the effort.

6-3a Modern Optical Telescopes

For most of the 20th century, astronomers faced a serious limitation on the size of astronomical telescopes. Telescope mirrors were made thick to avoid bending that would distort the reflecting surface, but those thick mirrors were heavy. The 5-m (200-in.) mirror on Mount Palomar weighs 14.5 tons. Those old-fashioned telescopes were massive and expensive. Today's

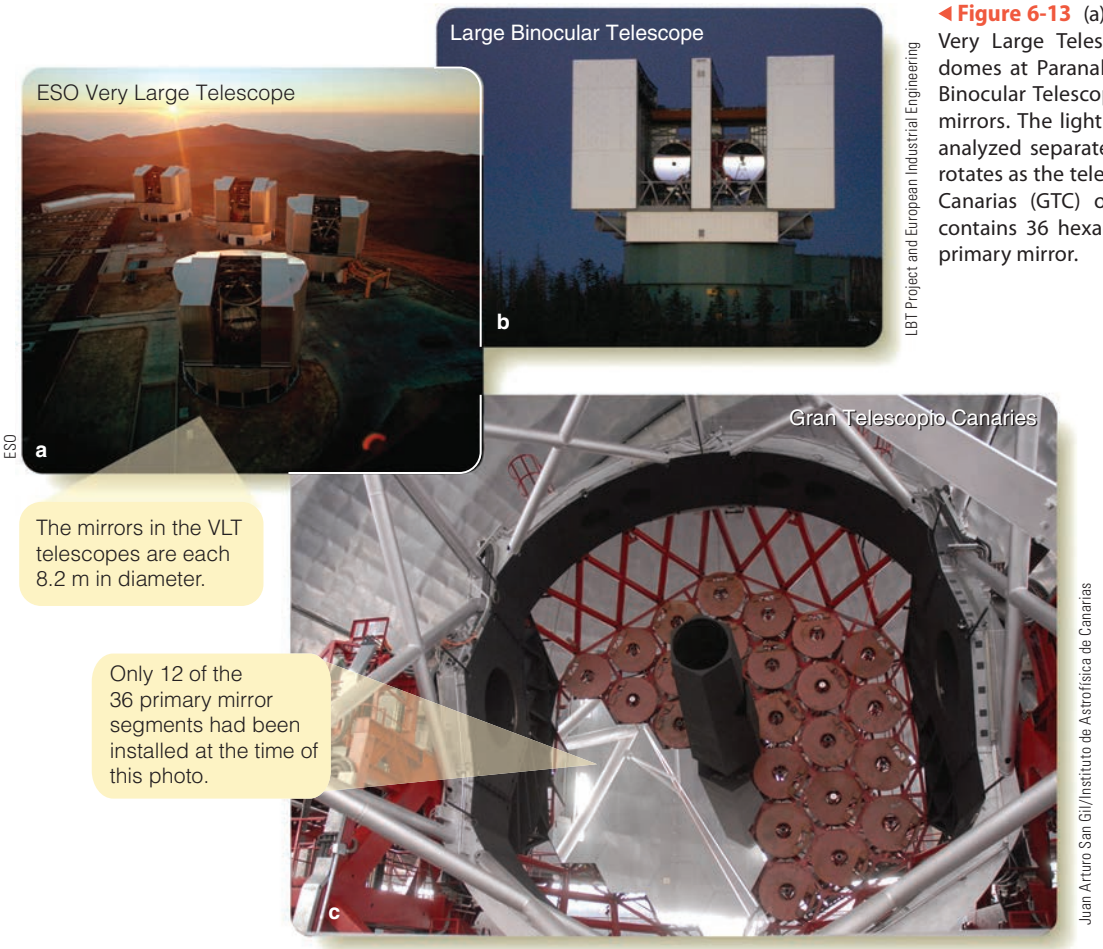
Richard Wainscoat, Aina Kai LLC

▲ **Figure 6-12** Aerial view of the optical, infrared, and radio telescopes on Mauna Kea in Hawai'i, 4200 m (nearly 14,000 ft) above sea level. The high altitude, low atmospheric moisture, lack of nearby large cities, and location near the equator make this mountain one of the best places on Earth to build an observatory.

astronomers have solved these problems in a number of ways. Look at **Modern Optical Telescopes** and notice three important points about telescope design and ten new terms that describe optical telescopes and their operation:

1. Conventional-design reflecting telescopes use large, solid, heavy mirrors to focus starlight to a *prime focus,* or by using a *secondary mirror,* to a *Cassegrain focus* (pronounced *KASS-uh-grain*). Other telescopes have a *Newtonian focus* or a *Schmidt-Cassegrain focus.*
2. Telescopes must have a *sidereal drive* to follow the stars. An *equatorial mount* with motion around a *polar axis* is the conventional way to provide that motion. Today, astronomers can build simpler, lighter-weight telescopes on *alt-azimuth mounts* that depend on computers to move the telescope so that it follows the apparent motion of stars as Earth rotates without having an equatorial mount and polar axis.
3. *Active optics,* computer control of the shape of a telescope's main mirrors, allows the use of thin, lightweight mirrors—either "floppy" mirrors or segmented mirrors. Reducing the weight of the mirror reduces the weight of the rest of the telescope, making it stronger and less expensive. Also, thin mirrors cool and reach a stable shape faster at nightfall, producing better images during most of the night.

Modern engineering techniques and high-speed computers have allowed astronomers to build and use new, giant telescopes with unique designs. A few are shown in **Figure 6-13**. The European


ESO Very Large Telescope
a
Large Binocular Telescope
b
Gran Telescopio Canarias
c
The mirrors in the VLT telescopes are each 8.2 m in diameter.
Only 12 of the 36 primary mirror segments had been installed at the time of this photo.

ESO

LBT Project and European Industrial Engineering

Juan Arturo San Gil/Instituto de Astrofísica de Canarias

◀ **Figure 6-13** (a) The four telescopes of the European Very Large Telescope (VLT) are housed in separate domes at Paranal Observatory in Chile. (b) The Large Binocular Telescope (LBT) in Arizona carries two 8.4-m mirrors. The light gathered by the two mirrors can be analyzed separately or combined. The entire building rotates as the telescope moves. (c) The Gran Telescopio Canarias (GTC) on La Palma in the Canary Islands contains 36 hexagonal mirror segments in its 10.4-m primary mirror.

Modern Optical Telescopes

1 Reflecting telescopes with standard designs depicted on this page have capabilities limited by complexity, weight, and turbulence in Earth's atmosphere. Modern design solutions are shown on the opposite page.

The primary mirror makes light converge to a **prime focus** position high in the telescope tube, as shown at the right. Although the prime focus is a good place to image faint objects, it is inconvenient for large instruments. A **secondary mirror** can reflect the light through a hole in the primary mirror to a **Cassegrain focus**. This focal arrangement is the most common one for large telescopes.

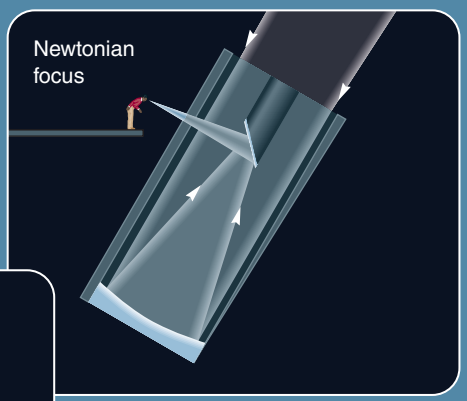
Secondary mirror

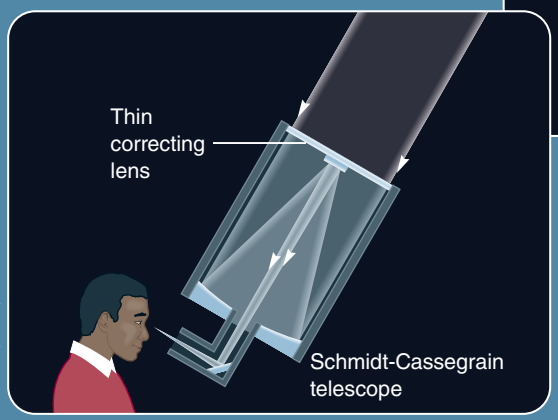
With large enough telescopes, astronomers can actually ride inside a prime-focus "cage," although observations are usually made using instruments connected to computers in a separate control room.

Conventional primary mirrors are thick to prevent the optical surface from sagging and distorting the image as the telescope is moved around the sky. But, large mirrors can weigh many tons, are difficult to support, and are expensive to make. Also, large mirrors take a long time to cool slowly after nightfall. Changes in shape as the mirror cools down make the telescope difficult to focus and cause image distortions.

The Cassegrain focus is convenient to access and has room for large, heavy instruments.

1a Smaller telescopes are often built with a **Newtonian focus**, the arrangement that Isaac Newton used in his first reflecting telescope. The Newtonian focus is inconvenient for large telescopes, as shown at right.

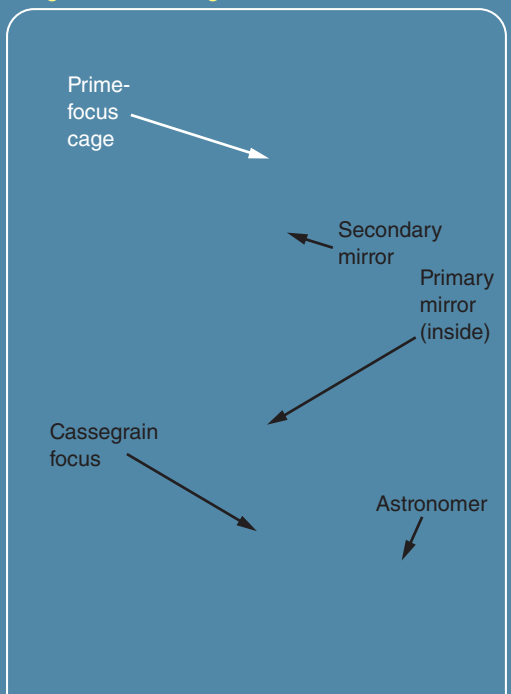




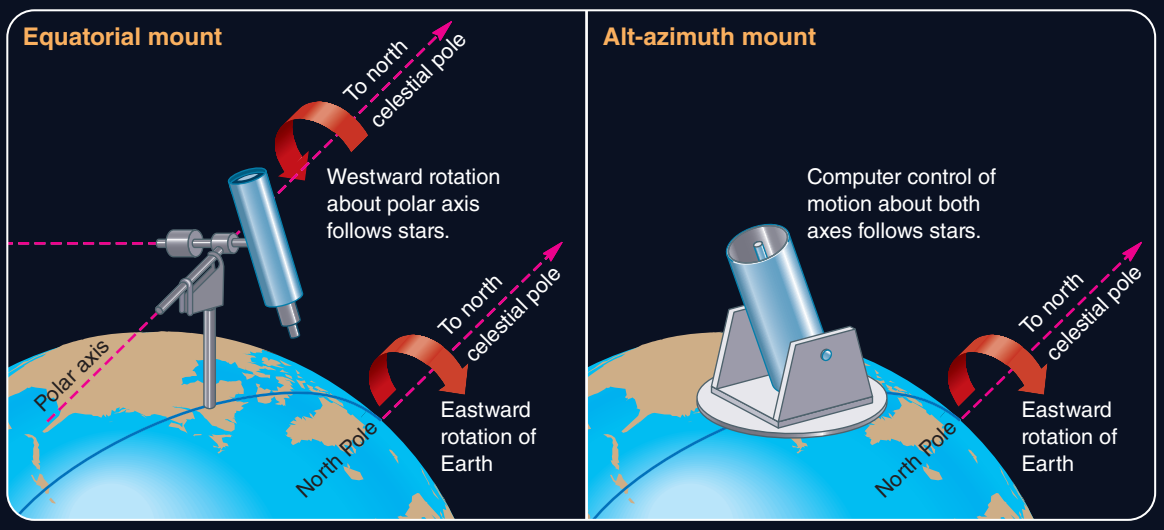
NOAO/AURA/NSF

1b Many small telescopes such as the one on the left use a **Schmidt-Cassegrain** focus. A thin correcting plate improves the image but is not curved enough to introduce serious chromatic aberration.

1c Shown below, observations using the 4-m Mayall Telescope at Kitt Peak National Observatory in Arizona can be made at either the prime focus or the Cassegrain focus. Note the human figure at lower right.

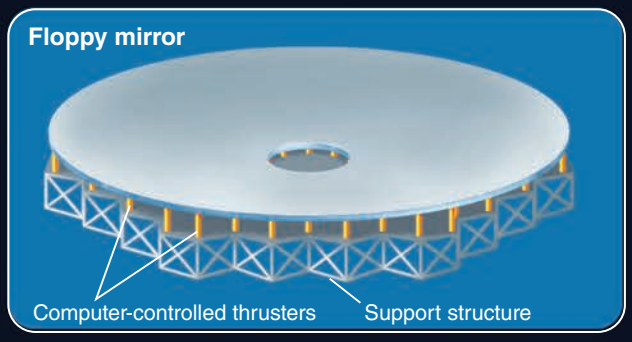


NOAO/AURA/NSF



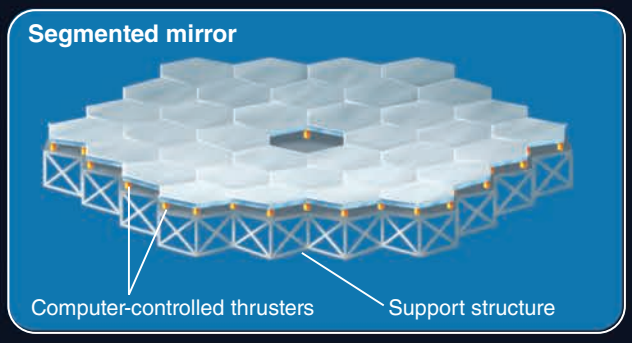
2 Telescope mountings must contain a **sidereal drive** to move the telescope smoothly westward, countering the eastward rotation of Earth. The earlier **equatorial mount** (*far left*) has a **polar axis** parallel to Earth's axis, but the mount used for the largest modern telescopes is **alt-azimuth mount** (altitude-azimuth; near left) moves like a cannon—up and down, left and right. Alt-azimuth mountings are simpler to build than equatorial mountings but require computer control to follow the stars.

3 Unlike traditional thick mirrors, thin mirrors, sometimes called "floppy" mirrors as shown at right, weigh less and require less massive support structures. Also, they cool rapidly at nightfall and there is less distortion from uneven expansion and contraction.



3a Grinding a large mirror may remove tons of glass and take months, but new techniques speed the process. Some large mirrors are cast in a rotating oven that causes the molten glass to flow to form a concave upper surface. Grinding and polishing such a preformed mirror is much less time consuming.

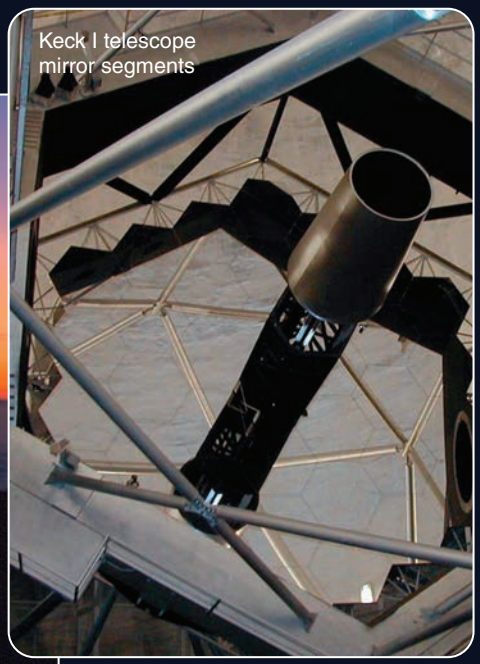
3b Mirrors made of segments are economical because the segments can be made separately. The resulting mirror weighs less and cools rapidly.



3c Both floppy mirrors and segmented mirrors sag under their own weight. Their optical shapes must be controlled by computer-driven thrusters behind the mirrors, a technique called **active optics**.

3d The two Keck telescopes, each 10 meters in diameter, are located atop the extinct volcano Mauna Kea in Hawai'i. Their two primary mirrors are composed of hexagonal mirror segments, as shown at right.

W.M. Keck Observatory

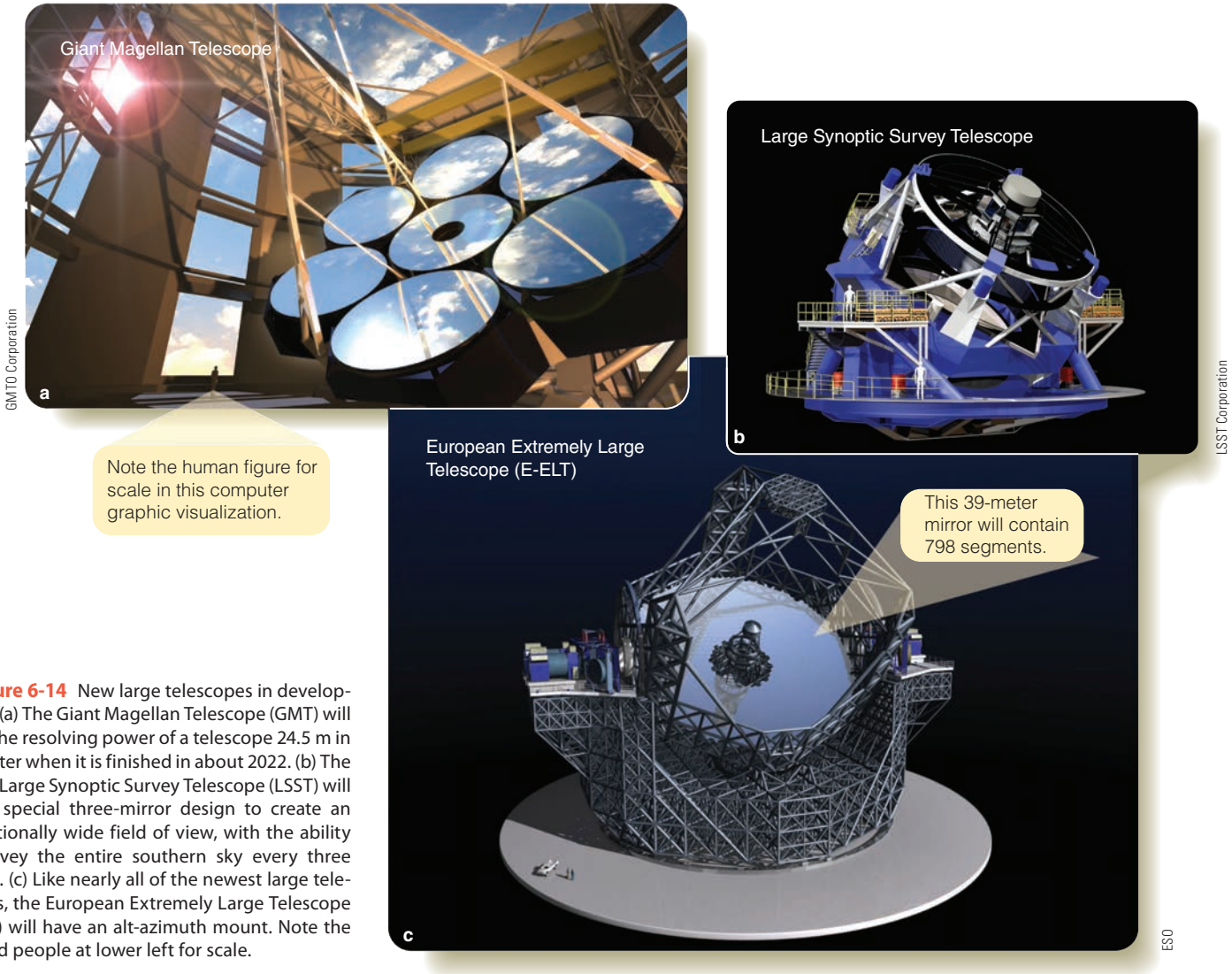


Southern Observatory built the Very Large Telescope (VLT) in the foothills of the Andes Mountains in northern Chile. The VLT actually consists of four telescopes, each with a mirror 8.2 m (323 in., about 27 ft) in diameter and only 17.5 cm (6.9 in.) thick. U.S. and Italian astronomers have built the Large Binocular Telescope (LBT) on Mount Graham in Arizona. The LBT carries a pair of 8.4-m (331-in.) mirrors on a single mount. The twin Keck telescopes on Mauna Kea in Hawai'i have primary mirrors 10 m (394 in.) in diameter that are each made of 36 individually controlled hexagonal segments. The Gran Telescopio Canarias (GTC), located atop a volcanic peak in the Canary Islands, carries a segmented mirror 10.4 m (410 in., over 34 ft) in diameter and is, at the time of this writing, the largest single telescope in the world.

Other giant telescopes are being planned for completion in the 2020s, all with segmented or multiple mirrors (**Figure 6-14**). The Giant Magellan Telescope (GMT) will carry seven asymmetrically curved thin mirrors, each 8.4 m in diameter, on a single mounting. It will be located in Chile and have the light-gathering power of a single 22.0-m telescope. The Thirty Meter Telescope (TMT), under development by an international team led by U.S. astronomers, is planned to have a mirror up to 30 meters (100 ft) in diameter comprised of 492 hexagonal segments and was intended for a site on Mauna Kea in Hawai'i. In 2016, TMT construction was halted by the Hawai'i Supreme Court because of a petition filed by native Hawaiian activists and allied groups opposed to the project. At the time of this writing in early 2017, the TMT directors were beginning a search for alternate sites.

Another international team is designing the European Extremely Large Telescope (E-ELT) to carry 906 segments, making up a mirror 39 meters in diameter . The E-ELT will be built on Cerro Armazones, a mountain in the Atacama Desert of Chile. Other very large telescopes have been proposed with estimated completion dates of 2020 or later.

A ground-based telescope is normally operated by astronomers and technicians working in a control room in the same building, but some telescopes are now used by astronomers many miles, even thousands of miles, from the observatory. Other telescopes are fully automated and operate without direct human supervision. That, plus continuous improvement in computer speed and storage capacity, has made possible huge surveys of the sky in which millions of objects have been observed or are planned for observation. For example, the Sloan Digital Sky Survey (SDSS)



▲ **Figure 6-14** New large telescopes in development. (a) The Giant Magellan Telescope (GMT) will have the resolving power of a telescope 24.5 m in diameter when it is finished in about 2022. (b) The 8.4-m Large Synoptic Survey Telescope (LSST) will use a special three-mirror design to create an exceptionally wide field of view, with the ability to survey the entire southern sky every three nights. (c) Like nearly all of the newest large telescopes, the European Extremely Large Telescope (E-ELT) will have an alt-azimuth mount. Note the car and people at lower left for scale.

mapped the entire Northern Hemisphere sky, measuring the position and brightness of 100 million stars and galaxies at five ultraviolet, optical, and infrared wavelengths.

The data from SDSS are available for you to examine and manipulate at this website: skyserver.sdss.org. Some of the SDSS data are also at Microsoft's World Wide Telescope (www.worldwidetelescope.org) and Google Sky (www.google.com/sky/). The "citizen science" Galaxy Zoo site (www.galaxyzoo.org) allows volunteers (this could mean you) to classify galaxies based on their appearances in SDSS images.

The future Large Synoptic Survey Telescope (LSST) has an 8.4-m primary mirror already completed; construction of facilities on Cerro Pachón in Chile are expected to be completed in 2019 (Figure 6-14b). Using a 3.2-billion-pixel charge-coupled device (CCD) camera, LSST will be able to record the brightness at selected ultraviolet, visual, and infrared wavelengths of every object in one hemisphere of the sky brighter than magnitude 24.5 every three nights. Astronomers and private citizens will be studying those data for decades to come.

6-3b Modern Radio Telescopes

The dish reflector of a radio telescope, like the mirror of a reflecting telescope, collects and focuses radiation. Although a radio telescope's dish may be tens or hundreds of meters in diameter, the receiver antenna may be as small as your hand. Its function is to absorb the radio energy collected by the dish. Because radio wavelengths are in the range of a few millimeters to a few tens of meters, the dish only needs to be shaped to that level of accuracy, much less smooth than a good optical mirror. In fact, wire mesh works well as a mirror for all but the shortest-wavelength radio waves.

For 53 years, from 1963 until 2016, the largest single radio dish in the word was the 305-m (1000-ft) Arecibo telescope, built into a mountain valley in Puerto Rico (**Figure 6-15**). Such a large dish can't be supported easily; the primary mirror is a thin metallic surface supported above the valley floor by cables attached near the rim, and the antenna platform hangs above the dish on cables from towers built on three mountain peaks around the valley's rim. By moving the antenna above the dish, radio astronomers can point the telescope at any object that passes within 20 degrees of the zenith as Earth rotates. Since completion in 1963, the Arecibo telescope has been an international center of radio astronomy research. A 500-m (1650-ft) diameter radio telescope named FAST, nicknamed Tianyan, "The Eye of Heaven," began operating in 2016. It was built by the Chinese government in a mountain-ringed bowl-shaped valley like the one in Puerto Rico. FAST has more than 2.5 times the collecting area of Arecibo's dish.

A radio astronomer works under two disadvantages relative to optical astronomers: poor resolution and low signal intensity. Recall that the resolving power of a telescope depends on the diameter of the primary lens or mirror but also on the wavelength of the radiation. At very long wavelengths like those of radio waves, the diffraction fringes are quite large. This means that images or maps from individual radio telescopes generally don't show such fine details as are seen in optical images.

David Parker/Science Source

▲ **Figure 6-15** The 305-m (1000-ft) radio telescope in Arecibo, Puerto Rico, is nestled in a naturally bowl-shaped valley. The receiver platform is suspended over the dish. A consortium led by SRI International and Universities Space Research Association (USRA) manages Arecibo Observatory for the National Science Foundation (NSF).

The second handicap radio astronomers face is the low intensity of the radio signals. You learned previously that the energy of a photon depends on its wavelength. Photons of radio energy have such long wavelengths that their individual energies are quite low. The cosmic radio signals arriving on Earth are astonishingly weak—billions of times weaker than the strength of the signal from a commercial radio station. To get detectable signals focused on the antenna, radio astronomers must build large collecting areas either as single large dishes or by combining arrays of smaller dishes. Even then, because the radio energy from celestial objects is so weak, it must be strongly amplified before it can be measured and recorded.

Practicing SCIENCE

Why do astronomers build optical observatories at the tops of mountains? Precise and accurate measurements are so fundamental to doing science that scientists often take extreme steps to get them.

It is certainly not easy to build a large, complicated, and fragile optical telescope at the top of a high mountain, but it is worth the effort. A telescope on top of a high mountain is above the densest part of Earth's atmosphere, so there is less air to dim the incoming light. Even more important, the turbulence of thin air on a mountaintop is less able to disturb light waves than that of thick air, so the seeing is better. The resolving power of a large optical telescope on Earth's surface is set by atmospheric seeing rather than by the telescope's diffraction. It really is worth the trouble to build telescopes on the peaks of high mountains.

Astronomers also put radio observatories in special locations. ***What considerations might astronomers make in choosing the location for a new radio telescope?***

6-4 Airborne and Space Observatories

Ground-based telescope performance is limited by Earth's atmospheric turbulence and transparency. There are sophisticated techniques that partly compensate for atmospheric seeing, but a telescope in space has no such problem, and its resolution is defined only by diffraction.

Also, as you learned earlier in this chapter, a telescope on the ground must look through one of the open atmospheric "windows" (wavelength ranges). Most types of electromagnetic radiation arriving here from the Universe—gamma-rays, X-rays, ultraviolet, and much of the infrared—do not reach Earth's surface because they are partly or completely absorbed by Earth's atmosphere. To gather light with those blocked wavelengths, telescopes must go to high altitudes or into space. As you will learn in the next few chapters, objects that are cooler than stars, such as stars that are forming, produce lots of infrared and microwave radiation but relatively little visible or ultraviolet light. In contrast, cosmic catastrophes such as exploding stars make mostly gamma-rays and X-rays. Combining information from as wide a variety of wavelengths as possible allows astronomers to gain a more comprehensive understanding of the Universe.

6-4a Airborne Telescopes

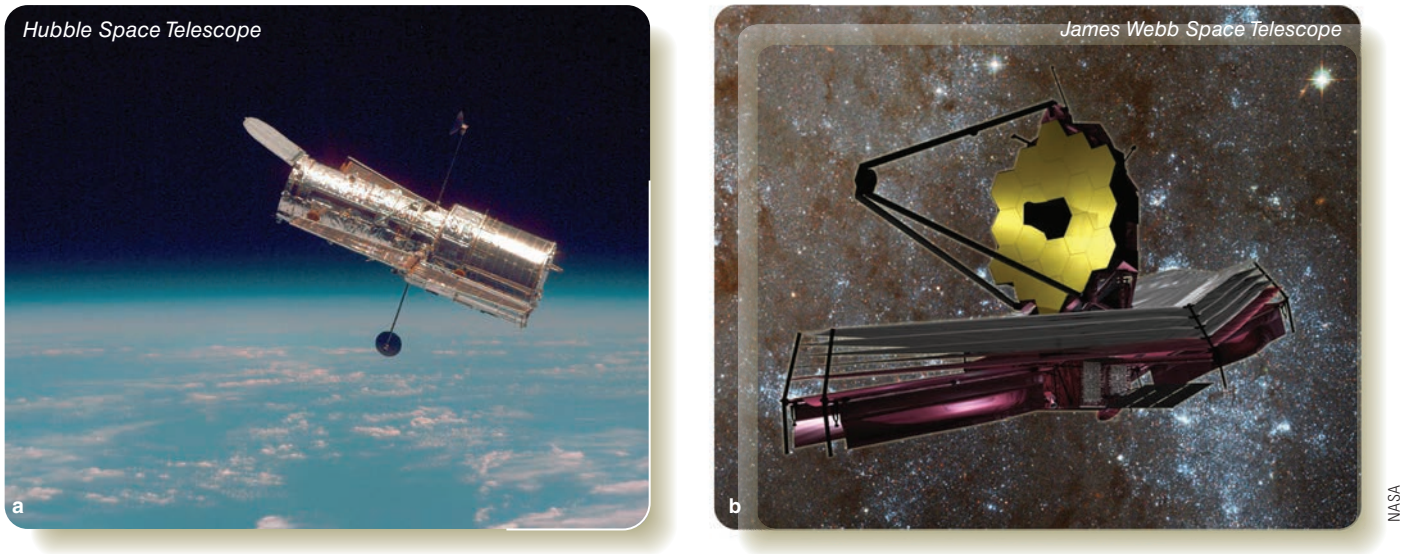
In addition to the atmospheric windows at visual and radio wavelengths you have already learned about, there are also a few narrow windows at short infrared wavelengths accessible from the ground, especially from high mountains such as Mauna Kea (Figure 6-12). However, most infrared wavelengths are blocked, especially by water vapor absorption. Also, Earth's atmosphere itself produces a strong infrared "glow." Observations at very long infrared wavelengths can only be made using telescopes carried to high altitudes by aircraft or balloons or launched entirely out of the atmosphere onboard spacecraft. (Notice that the reasons to put an infrared telescope above the atmosphere are not the same as the reasons to send an optical telescope into space.)

Starting in the 1960s, NASA developed a series of infrared observatories with telescopes carried above Earth's atmospheric water vapor by jet aircraft. Such airborne observatories are also able to fly to remote parts of Earth to monitor astronomical events not observable by any other telescope. The modern successor to those earlier flying observatories is the *Stratospheric Observatory for Infrared Astronomy,* or *SOFIA* (**Figure 6-16**). *SOFIA* consists of a 2.5-m (100-in.) telescope looking out an opening with a rollback door in the left side of a modified Boeing 747SP aircraft.

6-4b Space Telescopes

The most successful observatory in history, the *Hubble Space Telescope* (**Figure 6-17a**), is named after Edwin Hubble, the astronomer who discovered the expansion of the Universe. The *Hubble* telescope, also known as *HST,* was launched in 1990 and contains a 2.4-m (95-in.) mirror plus three instruments with which it can observe visible light plus some ultraviolet and infrared wavelengths. Its greatest advantage is the lack of seeing distortion, located completely above Earth's atmosphere. *Hubble* therefore can detect fine detail, and because it concentrates light into sharp images, it can detect extremely faint objects. It is

▲ **Figure 6-16** (a) The *Stratospheric Observatory for Infrared Astronomy (SOFIA),* a joint project of NASA and the German Aerospace Center (DLR), flies at altitudes up to 14 km (45,000 ft) where it can collect infrared radiation with wavelengths that are unobservable even from high mountaintops. (b) A visual-wavelength image of the planet Jupiter (*left*) compared with a composite infrared image (*right*) using images at wavelengths of 5.4, 24, and 37 microns made during *SOFIA*'s "First Light" flight in 2010. The white stripe in the infrared image is a region of relatively transparent clouds through which the warm interior of the planet can be seen.



▲ **Figure 6-17** (a) The *Hubble Space Telescope (HST)* orbits Earth at an average altitude of 545 km (340 mi) above the surface. In this image, the telescope is viewing toward the upper left. (b) Artist's conception of *HST's* eventual successor, the *James Webb Space Telescope (JWST). JWST* will be located in solar orbit almost 1 million miles from Earth, four times as far away as the Moon. It will not have an enclosing tube, thus resembling a radio dish more than a conventional optical telescope. *JWST* will observe the Universe from behind a multilayered sunscreen larger than a tennis court.

controlled from a research center on Earth and observes almost continuously. Nevertheless, the telescope has time to complete only a fraction of the many projects proposed by astronomers from around the world.

Hubble has been visited a number of times by the space shuttle so that astronauts could service its components and install new cameras and other instruments. Thanks to the work of the space shuttle crew who visited in 2009 and accomplished another refurbishment of the telescope's instruments, batteries, and gyroscopes, *Hubble* will almost certainly last until it can be replaced by the *James Webb Space Telescope (JWST),* which is expected to be ready for launch in 2018. *JWST* telescope will be launched into a solar orbit to avoid interference from Earth's strong infrared glow. Its primary mirror is a cluster of beryllium mirror segments that will open in space to form a 6.5-m (256-in.) mirror (**Figure 6-17b**).

Telescopes carrying long-wavelength infrared detectors must carry coolant such as liquid helium to chill their optics to near absolute zero temperature (−273°C or −460°F) so that heat radiation from the insides of the telescope and instruments does not blind the detectors. Such observatories have limited lifetimes because the coolant eventually runs out. The European Space Agency's *Herschel* 3-meter infrared space telescope, named after the scientist who discovered infrared radiation (Figure 6-2), was launched into solar orbit in 2009 together with the smaller *Planck* space observatory that studied millimeter-wavelength radiation. *Herschel* and *Planck* made important discoveries concerning distant galaxies, star formation, planets orbiting other stars, and the origin of the Universe during their 4-year lifetimes.

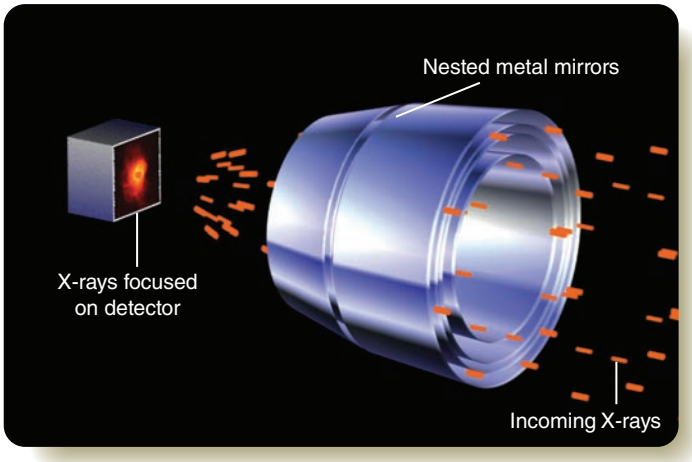
In 2016, NASA announced plans to build and launch a near-infrared space observatory named the *Wide-Field Infrared Survey Telescope*, or *WFIRST*. It is be designed for a dual mission: mapping the shapes and distributions of distant galaxies to understand the formation of large-scale structures in the early Universe and detecting and studying planets orbiting stars in the Milky Way Galaxy.

6-4c High-Energy Astronomy

Like infrared-emitting objects, gamma-ray, X-ray, and ultraviolet sources in the Universe are difficult to observe because the telescopes must be located high in Earth's atmosphere or in space. Also, high-energy photons are difficult to bring to a focus.

The first high-energy astronomy satellite, *Ariel 1,* was launched by the United Kingdom in 1962 and made solar observations in the ultraviolet and X-ray segments of the electromagnetic spectrum. Since then, many more space telescopes have followed *Ariel*'s lead. Some high-energy astronomy satellites such as *XMM-Newton,* an X-ray observatory developed by a consortium of European and British astronomers, have been general-purpose telescopes that observe many different kinds of objects. In contrast, some space telescopes are designed to study a single question or a single object. For example, the Japanese satellite *Hinode* (pronounced, *hee-no-day*) studies the Sun continuously at visual, ultraviolet, and X-ray wavelengths, and the *Kepler* space observatory has operated for more than 8 years detecting planets orbiting stars other than the Sun.

The largest X-ray telescope to date is the *Chandra X-ray Observatory (CXO). Chandra* operates in an orbit that extends a third of the way to the Moon so that it spends 85 percent of the time above the belts of charged particles surrounding Earth that



NASA/SAO/CXC

▲ **Figure 6-18** X-rays that hit a mirror at grazing angles are reflected like a pebble skipping across a pond. Thus, X-ray telescope mirrors like the ones in *Chandra* are shaped like barrels rather than dishes.

would produce electronic noise in its detectors. (*Chandra* is named for the late Indian American Nobel laureate Subrahmanyan Chandrasekhar, who was a pioneer in many branches of theoretical astronomy.) Focusing X-rays is difficult because they penetrate into most mirrors, so astronomers devised cylindrical mirrors in which the X-rays reflect at shallow angles from the polished inside of the cylinders to form images on X-ray detectors, as shown in **Figure 6-18**. The *Chandra* observatory has made important discoveries about everything from star formation to monster black holes in distant galaxies that will be described in later chapters.

The first large gamma-ray space telescope was the *Compton Gamma Ray Observatory,* launched in 1991. It mapped the entire sky at gamma-ray wavelengths. The European-built *INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL)* satellite was launched in 2002 and continues to be very productive in the study of violent eruptions of stars and black holes. The *Fermi Gamma-ray Space Telescope (FGST),* launched in 2008 and operated by a consortium of nations led by the United States, is capable of making highly sensitive gamma-ray maps of large areas of the sky.

Modern astronomy has come to depend on observations that cover the entire electromagnetic spectrum. More orbiting space telescopes are planned that will be even more versatile and sensitive than the ones operating now.

6-5 Astronomical Instruments and Techniques

Just looking through a telescope doesn't tell you much. A star looks like a point of light. A planet looks like a little disk. A galaxy looks like a hazy patch. To use a research telescope to learn about the Universe, you need to carefully analyze the light the telescope gathers. Special instruments attached to the telescope make that possible.

6-5a Cameras and Photometers

The **photographic plate** was the first device used by astronomers to record images of celestial objects. Photographic plates can detect faint objects in long time exposures and can be stored for later analysis. Brightness of objects imaged on a photographic plate can be measured with a lot of hard work that yields only moderate precision. Astronomers also build **photometers**, sensitive light meters used to measure the brightness of individual objects very precisely.

Present-day astronomers use **charge-coupled devices (CCDs)** as both image-recording devices and photometers. A CCD is a specialized computer chip containing millions of microscopic light detectors arranged in an array as small as a postage stamp. CCD chips have replaced photographic plates because they have some important advantages. CCDs are much more sensitive than photographic plates and can detect both bright and faint objects in a single exposure. Also, CCD images are digitized, meaning converted to numerical data, and thus can be stored in a computer's memory for later analysis. Although astronomy research-grade CCDs are extremely sensitive and therefore expensive, less sophisticated CCDs are now part of everyday life. You are familiar with them in digital cameras (both still and video) as well as in cell phone cameras.

Infrared astronomers use **array detectors** that are similar in operation to optical CCDs. At other wavelengths, photometers are still used for measuring brightness of celestial objects. Array detectors and photometers generally must be cooled to operate properly (**Figure 6-19**).

The digital data representing an image from a CCD or other array detector are easy to manipulate to bring out details that would not otherwise be visible. For example, astronomical

Adding liquid nitrogen to the camera on a telescope is a familiar task for astronomers.

Kris Koenig/Coast Learning Systems

▲ **Figure 6-19** Astronomical cameras with CCD and other types of array detectors must be cooled to low temperatures to operate properly, and that is especially true for infrared cameras.

images are often reproduced as negatives, with the sky white and the stars dark. That makes the faint parts of the image easier to see (Figure 6-20). Astronomers also can manipulate data to produce **false-color images** (or *representational-color*) in which the colors represent different aspects of the object such as intensity, rather than visual color. For example, because humans can't see radio waves, astronomers must convert radio data into something perceptible. One way is to measure the strength of the radio signal at various places in the sky and produce a false-color map in which each color marks areas of similar radio intensity. You can compare such a map to a weather map in which the different colors mark areas forecast to have different types and amounts of precipitation (Figure 6-21a). False-color images and maps are very commonly used in nonoptical astronomy (Figures 6-21b and 6-21c).

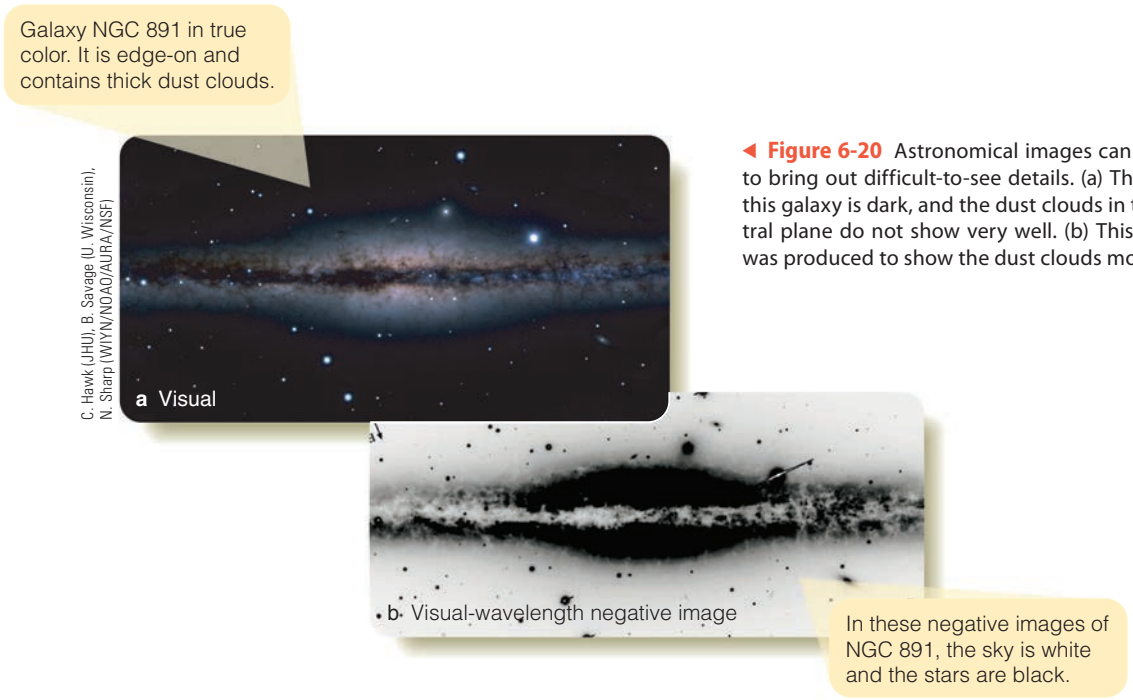
6-5b Spectrographs

To analyze light in detail, astronomers spread out the light according to wavelength (color), a function performed by a **spectrograph**. You can understand how this instrument works if you imagine repeating an experiment performed by Isaac Newton in 1666. Newton bored a small hole in the window shutter of his room to admit a thin beam of sunlight. When he placed a prism in the beam, it spread the light into a beautiful spectrum that splashed across his wall. From that and related experiments, Newton concluded that white light is made of a mixture of all the colors.

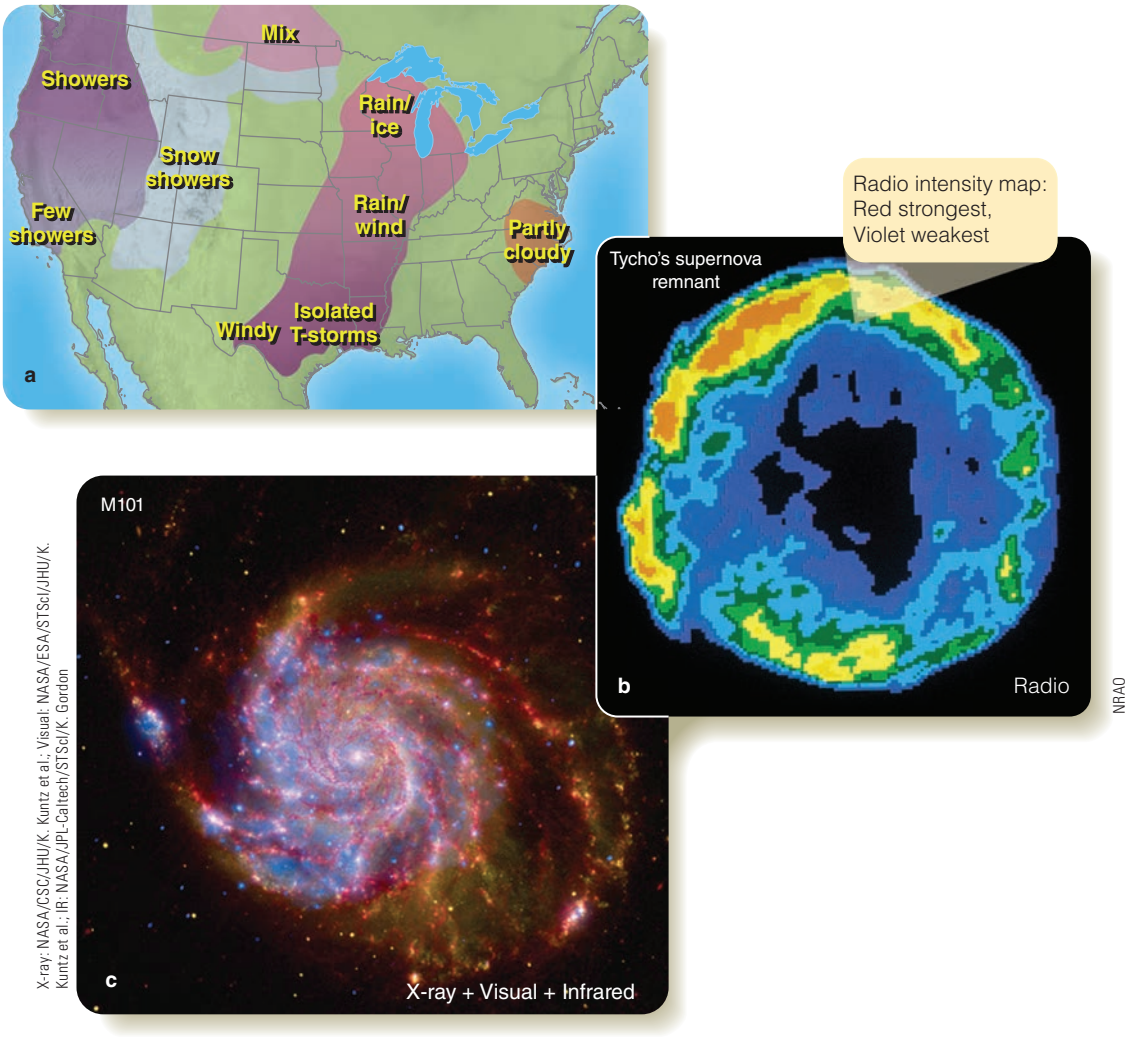
As you learned previously in regard to chromatic aberration of refracting telescopes, light passing from one medium such as air into another medium such as glass has its path bent at an angle that depends on its wavelength. For example, blue (shortwavelength) light passing through a prism bends the most, and red (long-wavelength) light bends least. Thus, the white light entering the prism is spread into a spectrum exiting the prism (Figure 6-22). You can build a simple spectrograph by using a narrow opening to define the incoming light beam, a prism to spread the light into its component colors, and a lens to guide the light into a camera.

Almost all modern spectrographs use a **grating** rather than a prism. A grating is a piece of glass or metal with thousands of parallel microscopic grooves scribed onto its surface. Different wavelengths of light reflect from or pass through the grating at slightly different angles, so white light encountering the grating is spread into a spectrum. You have probably noticed this effect when you look at the closely spaced lines etched onto a CD or DVD: As you tip the disk, different colors flash across its surface. A modern spectrograph can be built using a high-quality grating to separate light by wavelength, plus a CCD detector to record the resulting spectrum.

You will learn in the next chapter that spectra of astronomical objects such as stars and planets usually contain **spectral lines**—dark or bright lines that occur in spectra at specific wavelengths. Spectral lines are produced by atoms and molecules in the atmospheric gases of those objects. To measure the precise wavelengths of individual spectral lines and identify the atoms that produced them, astronomers use a **comparison spectrum**. Special light bulbs in the spectrograph produce bright spectral lines, or cells of gas in the spectrograph add dark lines, that are recorded next to the unknown spectrum. The wavelengths of the comparison spectral lines have been measured to high precision in laboratories, so astronomers can use comparison spectra as


Galaxy NGC 891 in true color. It is edge-on and contains thick dust clouds.
C. Hawk (JHU), B. Savage (U. Wisconsin), N. Sharp (WIYN/NOAO/AURA/NSF)
a Visual
b Visual-wavelength negative image
In these negative images of NGC 891, the sky is white and the stars are black.

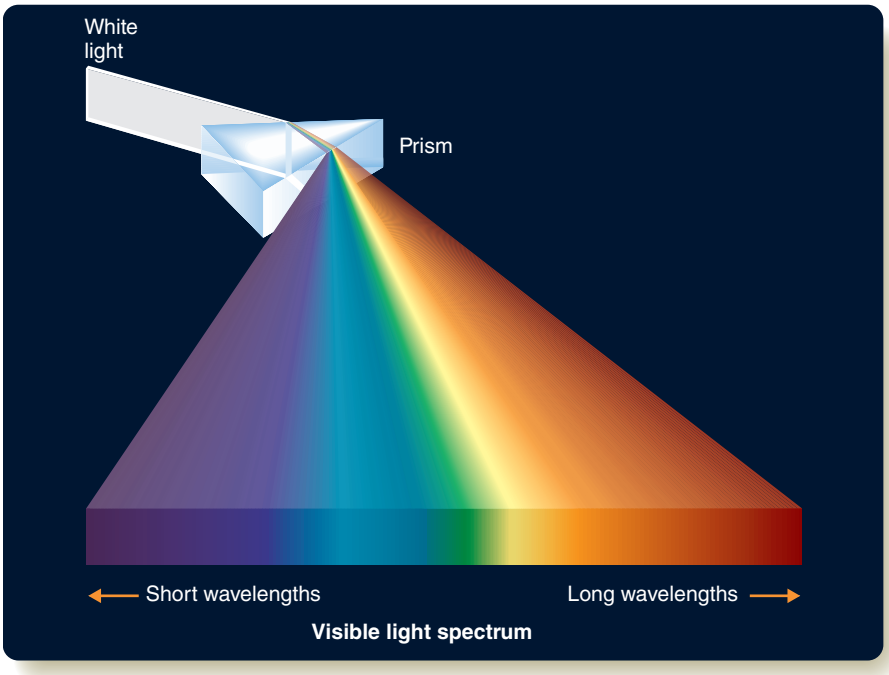
◀ **Figure 6-20** Astronomical images can be manipulated to bring out difficult-to-see details. (a) The color photo of this galaxy is dark, and the dust clouds in the galaxy's central plane do not show very well. (b) This negative image was produced to show the dust clouds more clearly.



X-ray: NASA/CSC/JHU/K. Kuntz et al.; Visual: NASA/ESA/STScI/JHU/K. Kuntz et al.; IR: NASA/JPL-Caltech/STScI/K. Gordon

NRAO

◀ **Figure 6-21** (a) A typical weather map uses contours with added color to show which areas are likely to receive precipitation, and what type. (b) A radio image of Tycho's supernova remnant, the expanding shell of gas produced by the explosion of a star first seen on Earth in 1572. This image's false-color code shows intensity of radio radiation at just one wavelength. (c) A picture of galaxy M101 composed of a visual-wavelength image from the *Hubble Space Telescope* combined with an X-ray image made by the *Chandra X-ray Observatory* plus an infrared image from the *Spitzer Space Telescope*. In this false-color image, the blue shows X-rays from hot gas heated by exploding stars and black holes, whereas red shows infrared emission from cool, dusty clouds of gas in which stars are being born.



◀ **Figure 6-22** A prism bends light by an angle that depends on the wavelength of the light. Short wavelengths bend most and long wavelengths least. Thus, white light passing through a prism is spread into a spectrum.

standards to measure wavelengths and identify spectral lines in the spectra of stars, galaxies, or planets.

Because scientists understand the details of how light interacts with matter, a spectrum carries a tremendous amount of information. That makes a spectrograph the astronomer's most powerful instrument. In the next chapter, you will learn more about the information astronomers can extract from a spectrum. Some astronomers say, "We don't know anything about an object until we get a spectrum," and that is only a slight exaggeration.

6-5c Adaptive Optics

You have already learned about active optics, which is a technique to adjust the shape of telescope optics slowly, compensating for effects of changing temperature as well as gravity bending the mirror when the telescope points at different locations in the sky. **Adaptive optics** is a more sophisticated technique that uses high-speed computers to monitor the distortion produced by turbulence in Earth's atmosphere and rapidly alter some optical components to correct the telescope image, sharpening a fuzzy blob into a crisp picture. The resolution of the image is still limited by diffraction in the telescope, but removing much of the seeing distortion produces a dramatic improvement in the detail that is visible (Figure 6-23a).

To monitor the distortion in an image, adaptive optics systems must look at a fairly bright star in the field of view, but there is not always such a star conveniently located near a target object such as a faint galaxy. In that case, astronomers can point a laser in a direction very close to that of their target object (Figure 6-23b). The laser causes gas in Earth's upper atmosphere to glow, producing an artificial star called a **laser guide star** in the field of view. The adaptive optics system can use information from the changing shape of the artificial star's image to correct the image of the fainter target.

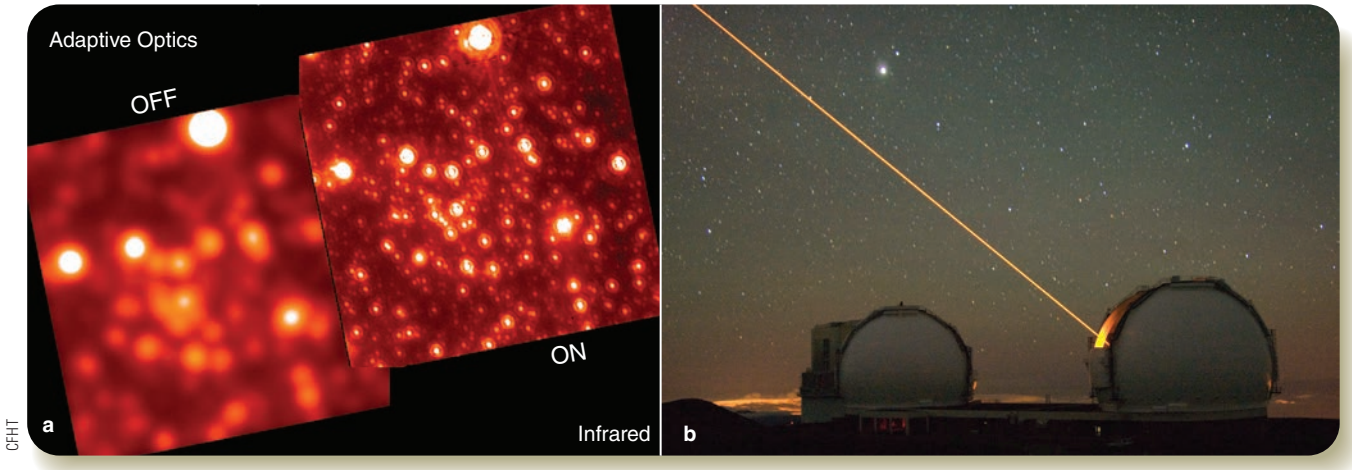
You have read about huge existing and planned optical telescopes 10 or more meters in diameter composed of segmented mirrors. Those telescopes would be much less useful without the addition of adaptive and active optics.

6-5d Interferometry

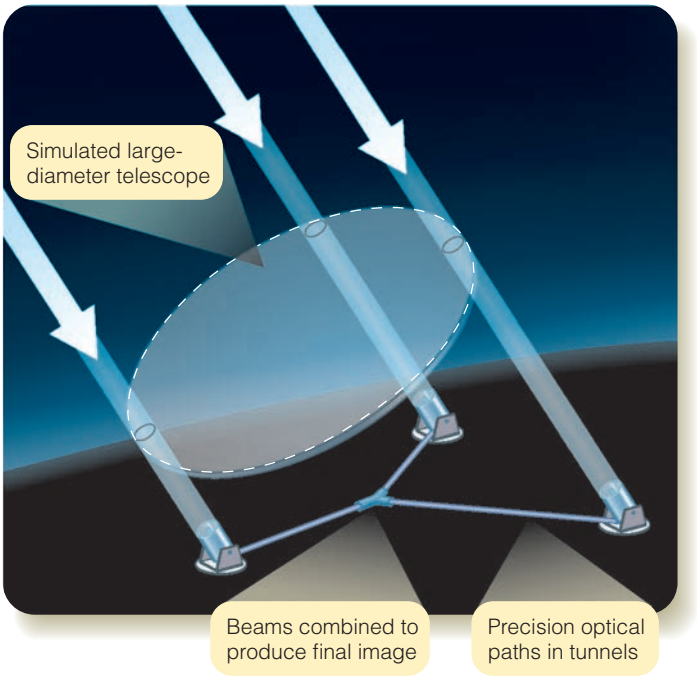
One of the reasons astronomers build big telescopes is to increase resolving power. Astronomers have been able to achieve very high resolution by connecting multiple telescopes together to work, in a sense, as if they comprised a single, very large telescope. This method of synthesizing a large "virtual" telescope from two or more smaller telescopes is known as **interferometry** (Figure 6-24). The images from such an interferometric telescope are not limited by the diffraction fringes of the individual small telescopes but rather by the diffraction fringes of the much larger virtual telescope.

In an interferometer, light from the separate telescopes must be brought together and combined carefully. The path that each light beam travels must be controlled so that it is known to a precision of a small fraction of the light's wavelength. Turbulence in Earth's atmosphere constantly distorts incoming light, so high-speed computers must continuously adjust the light paths.

As you already know, the resolving power of a radio telescope is relatively low. A dish 25 m in diameter receiving radiation with a wavelength of 21 cm has a resolution of only about 0.5 degree. In other words, a radio telescope 100 ft across is unable to detect any details in the sky smaller than the apparent size of the Moon. But, because long-wavelength radio waves are relatively easy to manipulate, radio astronomers were the first to learn how to combine two or more telescopes to form an interferometer capable of much higher resolution than a single telescope.



▲ **Figure 6-23** (a) In these images of the center of our galaxy, the adaptive optics system was turned "Off" for the left image and "On" for the right image. In the "On" image, the images of stars are sharper because the light is focused into smaller images; fainter stars are visible. (b) The laser beam shown leaving one of the Keck telescopes produces an artificial star in the field of view, and the adaptive optics system uses that laser guide star as a reference to reduce seeing distortion in the entire image.



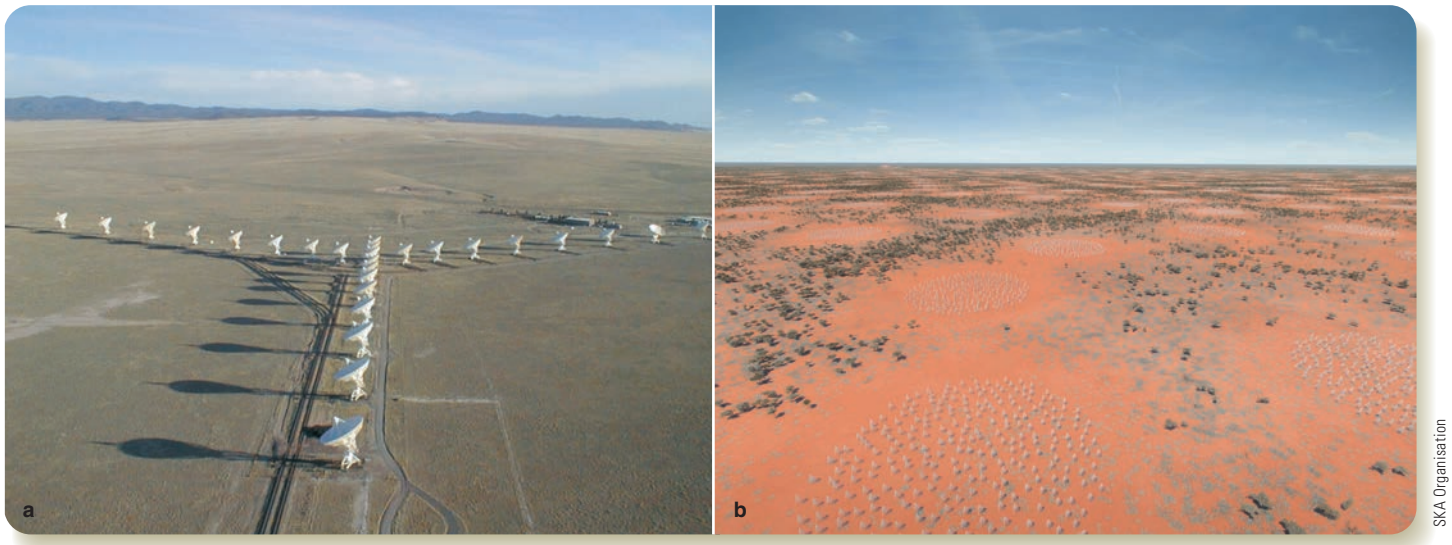
▲ **Figure 6-24** In an astronomical interferometer, smaller telescopes can combine their light to simulate a larger telescope with a resolution set by the separation of the smaller telescopes.

Radio interferometers must be quite large. The Very Large Array (VLA) consists of 27 dish antennas spread across the New Mexico desert (**Figure 6-25a**). In combination, they have the resolving power of a radio telescope up to 36 km (22 mi) in diameter. The VLA can resolve details smaller than 1 arc second, rivaling the performance of a large optical telescope at a good site. The Very Long Baseline Array (VLBA) that includes the VLA consists of matched radio dishes spread from Hawai'i to the Virgin Islands, with an effective diameter almost as large as Earth.

The Atacama Large Millimeter/submillimeter Array (ALMA) is an interferometric facility located on the Chajnantor plateau in northern Chile at an altitude of 5050 m (16,600 ft; see image on the first page of this chapter). It is described as the most powerful telescope ever built because of its combination of total mirror collecting area and high spatial resolution. ALMA began supporting research observations in 2011 when the planned array of 66 high-precision dish antennas was about half complete. Astronomers from the entire world will be able to use ALMA without having to travel to Chile; because of the extreme altitude of the facility, observations and data analyses will all be done over the Internet.

Radio astronomers are now planning the Square Kilometer Array (SKA) that will contain thousands of radio receivers with total collecting area of a square kilometer (1 million m^2, 5 times larger than the Tianyan dish) spread over a distance of 6500 km (4000 mi; **Figure 6-25b**). These giant radio interferometers depend on state-of-the-art computers to combine signals properly and create radio maps.

Recall that the wavelength of light is very short, roughly 0.0005 mm, so building optical interferometers is one of the most difficult technical problems that astronomers face, but the challenge has been met in several instances. The European VLT (Figure 6-13a) consists of four 8.2-m telescopes that can operate separately, but the light they collect, along with light from three 1.8-m telescopes on the same mountaintop, can be brought together through underground tunnels. The resulting optical interferometer, known as the VLTI, can provide the resolution



NRAO/AUI

SKA Organisation

▲ **Figure 6-25** (a) The Very Large Array (VLA) radio dishes in New Mexico can be moved to different positions along a Y-shaped set of tracks. They are shown here in their most compact arrangement. Signals from the dishes are combined to create very high-resolution radio maps of celestial objects. (b) Artist's conception of the proposed Square Kilometer Array (SKA) that will have concentrations of radio receivers in two clusters, one in South Africa and one in Australia, separated by 6500 km (4000 mi).

(but, of course, not the light-gathering power) of a telescope 200 m in diameter (660 ft, bigger than two football fields).

Astronomers using the VLTI in 2009 made an image of the red giant star T Leporis with a resolution of 0.004 arc second, equivalent to being able to discern a two-story house on the Moon. The Center for High Angular Resolution Astronomy (CHARA) telescope array on Mt. Wilson in Southern California combines six 1-m telescopes to create resolving power equivalent to a telescope 300 m (one-fifth of a mile) in diameter. Other facilities such as the two Keck 10-m telescopes in Hawai'i and the Large Binocular Telescope in Arizona also are capable of operating as interferometers. Although turbulence in Earth's atmosphere can be partially averaged out in an interferometer, astronomers are considering the possibility of putting interferometers in space to avoid atmospheric turbulence altogether.

6-6 Non-Electromagnetic Astronomy

This chapter is focused on how to collect and analyze electromagnetic radiation from space. Other types of energy also arrive here bearing information from the rest of the Universe and deserve at least a brief mention.

6-6a Particle Astronomy

Cosmic rays are subatomic particles traveling through space at tremendous velocities. Almost no cosmic rays reach the ground, but some of them smash into gas atoms in Earth's upper atmosphere, and fragments of those atoms shower down to the ground. Those secondary cosmic rays are passing through you as you read this sentence and will continue to do so throughout your life. Other types of particles from space interact weakly and seldom with Earth atoms, so huge detectors must be built to catch and count them. Detectors for some kinds of cosmic rays have been carried on balloons or launched into orbit, whereas others have been built deep underground where layers of rock filter out all but the most penetrating particles.

Astronomers are not yet sure what produces cosmic rays. Incoming particles that have electric charges have been deflected by electromagnetic forces as they traveled through our galaxy, which means astronomers can't easily tell where their original sources are located. Some lower-energy particles of various types are known to come from the Sun, and there are indications that at least a few high-energy cosmic rays are produced by the violent explosions of dying stars or supermassive black holes at the centers of galaxies. You will meet these exotic objects again in later chapters.

6-6b Gravity Wave Astronomy

You learned in Section 5-3 that Einstein's general theory of relativity predicts another type of non-electromagnetic signal, called gravity waves, which should be produced by any mass in the Universe that accelerates. Gravity waves are extremely weak and were not detected directly until 2015 when the LIGO (Laser Interferometer Gravitational Wave Observatory) facilities in Louisiana and Washington State measured the signal from the merger of two massive black holes in a distant galaxy. The Evolved Laser Interferometry Space Antenna (eLISA) is LIGO's planned space-based counterpart. These first "gravity telescopes" will open up an entirely new window on the Universe that, for the span of a century, was only theoretically predicted to exist.

WHAT ARE WE?

Curious

Telescopes are creations of curiosity. You look through a telescope to see more and to understand more. The unaided eye is a detector with limited sensitivity, and the history of astronomy is the history of bigger and better telescopes gathering more and more light to search for fainter and more distant objects.

The old saying "Curiosity killed the cat" is an insult to the cat and to curiosity. We humans are curious, and curiosity is a noble trait—the mark of an active, inquiring mind. At the base of human curiosity lies the fundamental question, "What are we?" Telescopes extend and amplify our senses, but they also allow us to extend and amplify our curiosity about our place in the Universe.

When people find out how something works, they say their curiosity is satisfied. Curiosity is an appetite like hunger or thirst, but it is an appetite for understanding. As astronomy expands our horizons and we learn about how distant stars and galaxies form and evolve, we feel satisfaction partly because we are learning about ourselves and about how we fit in the Universe. We are beginning to understand what we are.

Study and Review

SUMMARY

6-1 Radiation: Information from Space

- Visual light is the visible form of **electromagnetic radiation**, which is an electric and magnetic disturbance that transports energy at the speed of light *c*. The **wavelength (λ)** of light, or the distance between the peaks of a wave, is usually measured in **nanometers (nm)** (10^{-9} m) or **angstroms (Å)** (10^{-10} m). **Frequency (ν)** is the number of waves that pass a stationary point in 1 second. The wavelength (λ) of an electromagnetic wave times its frequency (ν) equals the speed of light *c*.

$$\lambda\nu = c \qquad \text{(Eq. 6-1)}$$

  A **photon** is a packet of light waves that can act as a particle or as a wave. The energy carried by a photon is proportional to its frequency and inversely proportional to its wavelength.
- A **spectrum** is a display of light that is viewed or recorded after being sorted in order of wavelength or frequency. In addition to visible light, the complete electromagnetic spectrum includes **gamma-rays**, **X-rays**, **ultraviolet (UV) radiation** at the higher frequencies, and **infrared (IR) radiation**, **microwaves**, and **radio waves** at the lower frequencies. Earth's atmosphere is transparent in some **atmospheric windows**: visible light, shorter-wavelength infrared, and short-wavelength radio.

6-2 Refracting Telescopes

- **Refracting telescopes** use a **primary lens** to bend and focus the light into an image. **Reflecting telescopes** use a **primary mirror** to focus the light. The image produced by the telescope's primary lens or mirror can be magnified by an **eyepiece**. Lenses and mirrors with short **focal lengths** must be strongly curved and are more expensive to grind to an accurate shape.
- Because of **chromatic aberration**, refracting telescopes cannot bring all colors to the same focus, resulting in color fringes around the images. An **achromatic lens** partially corrects for this, but such lenses are expensive and cannot be made much larger than about 1 m (40 in.) in diameter.
- Reflecting telescopes are easier to build and less expensive than refracting telescopes of the same diameter. Also, reflecting telescopes do not suffer from chromatic aberration. Most large **optical telescopes** and all **radio telescopes** are reflecting telescopes.
- **Light-gathering power** refers to the ability of a telescope to collect light. **Resolving power** refers to the ability of a telescope to reveal fine detail and is expressed by the equation:

$$\alpha \text{ (arc seconds)} = 2.06 \times 10^5 \left(\frac{\lambda}{D}\right) \qquad \text{(Eq. 6-2)}$$

  Diffraction fringes in an image, caused by the interaction of light waves with the telescope's apertures, limit the amount of detail that can be seen. **Magnifying power** is the ability of a telescope to make an object look bigger.

6-3 Observatories on Earth

- Astronomers build optical observatories on remote, high mountains for two reasons: (1) The air on top of a mountain is relatively steady, and the **seeing** is better. (2) Observatories are located far from cities to avoid **light pollution**. Astronomers also build radio telescopes remotely but more for the reason of avoiding interference from human-produced radio noise.
- In a reflecting telescope, light first comes to a focus at the **prime focus**, but a **secondary mirror** can direct light to other locations such as the **Cassegrain focus**. The **Newtonian focus** and **Schmidt-Cassegrain focus** are other focus locations used in some smaller telescopes.
- Because Earth rotates, telescopes must have a **sidereal drive** to remain pointed at celestial objects. An **equatorial mount** with a **polar axis** is the simplest way to accomplish this. An **alt-azimuth mount** can support a more massive telescope but requires computer control to compensate for Earth's rotation.
- Very large telescopes can be built with **active optics** to control the mirror's optical shape. These telescopes usually have either one large, thin, flexible mirror or a mirror broken into many small segments. Advantages include mirrors that weigh less, are easier to support, and cool faster at nightfall. A major disadvantage is that the optical shape needs to be adjusted gradually and continuously to maintain a good focus.

6-4 Airborne and Space Observatories

- The turbulence in Earth's atmosphere distorts and blurs images. Telescopes in orbit are above this seeing distortion and are limited only by diffraction in their optics. Earth's atmosphere absorbs gamma-ray, X-ray, ultraviolet, far-infrared, and microwave light. To observe at these wavelengths, telescopes must be located at high altitudes or in space.

6-5 Astronomical Instruments and Techniques

- Astronomers in the past used **photographic plates** to record images at the telescope and **photometers** to precisely measure the brightness of celestial objects. Modern electronic systems such as **charge-coupled devices (CCDs)** and other types of **array detectors** have replaced both photographic plates and photometers in most applications.
- Electronic detectors have the advantage that data from them are automatically digitized in numerical format and can be easily recorded and manipulated. Astronomical images in digital form can be computer-enhanced to produce **false-color images** that bring out subtle details.
- **Spectrographs** using prisms or a **grating** spread light out according to wavelength to form a spectrum, revealing hundreds of **spectral lines** produced by atoms and molecules in the object being studied. A **comparison spectrum** that contains lines of known wavelengths allows astronomers to measure the precise wavelengths of individual spectral lines produced by an astronomical object.
- **Adaptive optics** techniques involve measuring seeing distortions caused by turbulence in Earth's atmosphere, then partially canceling out those distortions by rapidly altering some of the telescope's optical components. In some facilities a powerful laser beam is used to produce an artificial **laser guide star** high in Earth's atmosphere that can be monitored by an adaptive optics system.
- **Interferometry** refers to the technique of connecting two or more separate telescopes to act as a single large telescope that has a resolution equivalent to that of a single telescope with a diameter

that is as large as the separation between the individual telescopes. The first working interferometers were composed of multiple radio telescopes.

6-6 Non-Electromagnetic Astronomy

- **Cosmic rays** are not electromagnetic radiation; they are subatomic particles such as electrons and protons traveling at nearly the speed of light, arriving from mostly unknown cosmic sources.

REVIEW QUESTIONS

1. Does light include radio waves? Explain your answer.
2. Why would you not include sound waves in the electromagnetic spectrum?
3. If the frequency of an electromagnetic wave increases, does the number of waves passing by you increase, decrease, or stay the same? Does the wavelength increase, decrease, or stay the same? Does the energy of the photon increase, decrease, or stay the same?
4. Compared to infrared radiation, does ultraviolet radiation have longer or shorter wavelengths? Does ultraviolet radiation have higher or lower energy per photon?
5. Does red light have a higher or lower energy than blue light? Does red light have higher or lower frequency than blue light? Does red light have a longer or shorter wavelength than blue light?
6. If you had limited funds to build a large telescope, which type would you choose, a refractor or a reflector? Why?
7. Why do nocturnal animals usually have large pupils in their eyes? How is that related to the way astronomical telescopes work?
8. If you were in a deep dark cave and you turned off or completely covered all light sources, then waited awhile for your eyes to "adjust," would you be able to see? Why or why not?
9. Why do optical astronomers often put their telescopes at the tops of mountains, whereas radio astronomers sometimes put their telescopes in deep valleys?
10. What advantage, if any, would radio astronomers have by building their telescopes at the tops of mountains?
11. What are the advantages of making a telescope mirror thin? What problems result?
12. Small telescopes are often advertised as "200 power" or "magnifies 200 times." How would you change these advertisements to market to astronomers?
13. Why do single-dish radio telescopes have poor resolving power compared to optical telescopes of the same diameter?
14. The Moon has no sustained atmosphere. What advantages would you have if you built an observatory on the lunar surface?
15. Why must telescopes observing at long infrared (that is, far-infrared) wavelengths be cooled to low temperatures?
16. What purpose do the colors in a false-color image or map serve?
17. What might you detect with an X-ray telescope that you could not detect with an infrared telescope?
18. If you were looking for exploding stars, in which wavelength band would you want to observe? Why?
19. How is the phenomenon of chromatic aberration related to how a prism spectrograph works?
20. What is the role of prisms and gratings in spectrographs?
21. How are active optics different from adaptive optics?
22. Why would radio astronomers build identical radio telescopes in many different places around the world?
23. Give an example of a cosmic ray. Are cosmic rays electromagnetic waves? Why or why not?
24. **How Do We Know?** How is the resolution of an astronomical image related to the precision of a measurement?

ACTIVE INQUIRY QUESTIONS

1. When you think of a "nice night," what does the sky look like? You have a friend who is an astronomer. What would your friend think of your "nice night"? Does that depend on what type of astronomer he or she is (in other words, what wavelength of electromagnetic radiation that friend prefers to collect and study)?
2. Consider the "atmospheric windows" of Earth's atmosphere (Figure 6-2b). How do these match against the biological responses of animals and plants here on Earth? What do you think is the reason for that?

PROBLEMS

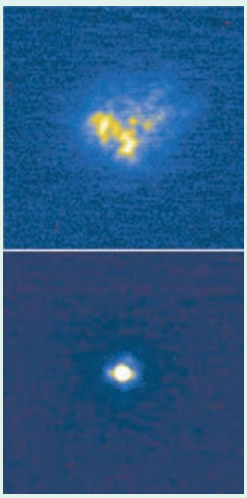
1. Plastic bags have a thickness about 0.001 mm. How many wavelengths of red light is that?
2. What is the wavelength of radio waves transmitted by a radio station with a frequency of 100 million cycles per second?
3. What is the frequency and wavelength of an FM radio station on your radio dial at 102.2?
4. Does a 700-nm photon have more or less energy than a 400-nm photon? By what factor more (or less)? What colors are associated with those wavelengths?
5. Compare the light-gathering power of a 10-m Keck telescope with that of a 0.6-m telescope.
6. How does the light-gathering power of one of the 10-m Keck telescopes compare with that of the human eye? Assume that the pupil of your eye can open to a diameter of about 0.8 cm in dark conditions.
7. Telescope A has a 60-in. diameter whereas Telescope B has a 4-cm diameter. Which telescope gathers more light and by how much?
8. What is the resolving power of Telescope A and Telescope B in Problem 7, at a wavelength of 550 nm? Which is better, and why?
9. In general, does a telescope resolve a close double star, such as in Figure 6-9b, better at blue wavelengths or red? How do you know?
10. What is the resolving power of a 25-cm (10-in.) telescope at a wavelength of 550 nm? What do two stars 1.5 arc seconds apart look like through this telescope?
11. Most of Galileo's telescopes were only about 2 cm in diameter. Would he have been able to resolve the two stars mentioned in Problem 10?
12. How does the resolving power of the Mount Palomar 5-m telescope compare with that of the 2.4-m *Hubble Space Telescope?* Why does *HST* generally still outperform the ground-based 5-m telescope?
13. If you build a telescope with a focal length of 1.3 m, what eyepiece focal length is needed for a magnification of 100 times?
14. Astronauts observing from a space station need a telescope with a resolving power of 0.1 arc second at a wavelength of 550 nm and a magnifying power of 250. Design a telescope to meet their needs. What will its light-gathering power be, compared with a dark-adapted human eye? (*Note*: See Problem 6).
15. A spy satellite orbiting 400 km above Earth is supposedly capable of counting individual people in a crowd in visual-wavelength images. Assume that the satellite's cameras operate at a wavelength of 550 nm. Assume an average person has a size of 0.7 m as seen from above. Estimate the minimum telescope diameter that the satellite must carry. (*Hint:* Use the small-angle formula [Eq. 3-1] to convert linear size to angular size.)

SENSE OF PROPORTION

1. (a) The wavelength of visible light is about 5000 times longer than the wavelength of typical X-rays. If you represent the wavelength of an X-ray photon by a line 1 in. long, how long a line (in feet) must you draw to represent the wavelength of visible light?
   (b) The wavelength of AM radio waves is about 6×10^8 larger than the wavelength of visible light. How long a line (in AU) must you add to your diagram to represent the wavelength of AM radio waves? (*Note*: Relevant unit conversions can be found in Appendix A.)
2. Spy satellites in orbit looking down at Earth can resolve objects as small as 1 ft in diameter; thus they can image individual human beings. When astronomers link near-infrared telescopes together to make an interferometer, they can resolve objects with about 400 times smaller in angular size. How small a detail could such a telescope image if it could look down at Earth from orbit? (Ignore the effects of atmospheric seeing.)

LEARNING TO LOOK

1. Estimate the wavelength of the water waves near the shoreline in Figure 6-1a.
2. How many atmospheric windows are shown in Figure 6-3b, and which bands of the electromagnetic spectrum are they in?
3. Locate the primary optical element in Figure 6-15. Is this a refracting telescope or a reflecting telescope? How can you tell?
4. Did the magnification, resolving, or light-gathering power change from the left image to the right image in Figure 6-23a? How do you know?
5. Explain what is meant by "intensity" in the single-wavelength false-color representation of Figure 6-21b. Would you have selected this false-color code pattern, or would you have selected red to represent a wavelength with low intensity?
6. These two images show a star before and after an adaptive optics system attached to the telescope was switched on. What causes the distortion in the first image, and how do adaptive optics improve the image?

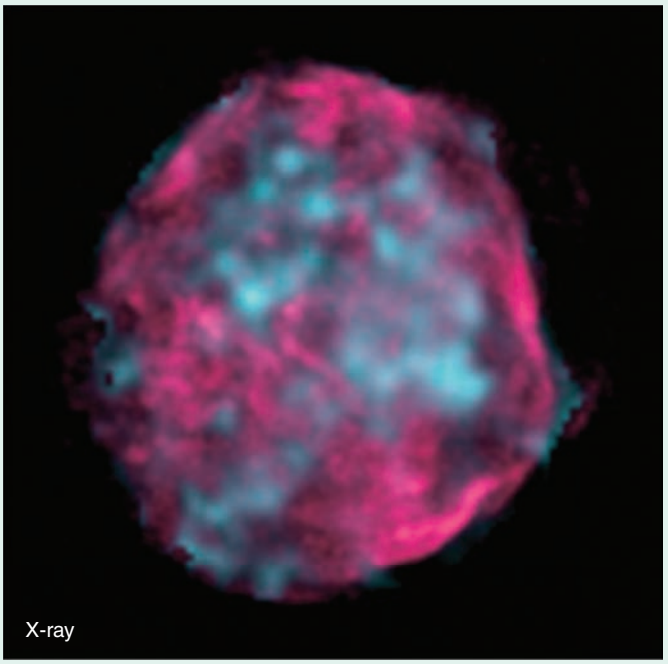


ESO

7. The star images in this photo are tiny disks, but the diameters of these disks are not related to the diameter of the stars. Explain why the telescope can't resolve the diameter of the stars. What causes the apparent diameters of the stars?

NASA/ESA/STScI/AURA/NSF/G. Meylan (École polytechnique fédérale de Lausanne)

8. This X-ray image shows the remains of an exploded star. Explain why images recorded by telescopes in space are often displayed in "false" color rather than in the "colors" (that is, wavelengths) received by the telescope. What color would we see this image if the image were not falsely colored?



NASA/CXC/S. Park (PSU)

7 Atoms and Spectra

RosaBetancourt 00 people images/
Alamy Stock Photo

▲ The world you see, including these neon signs, is animated by the properties of atoms and subatomic particles.

In the previous chapter, you read how telescopes gather light, cameras record images, and spectrographs spread light into spectra. Now you can consider why astronomers make such efforts. Here you will find answers to three important questions:

- How do atoms interact with light to produce spectra?
- What are the types of spectra that can be observed?
- What can be learned from spectra of celestial objects?

Up to this point, you have been considering what you can see with your eyes alone or aided by telescopes and astronomical instruments such as spectrographs. This chapter marks a change in the way you study nature: You will begin learning about the modern field of astrophysics that links physics experiments and theory to astronomical observations, and realize why the spectrum of an object can be so informative. In the chapters that follow, you will learn about the rich information derived from spectra of planets, stars, and galaxies that reveals the secrets of their internal structures and histories.

Awake! for Morning in the Bowl of Night
Has flung the Stone that puts the Stars to Flight:
And Lo! the Hunter of the East has caught
The Sultan's Turret in a Noose of Light.

THE RUBÁIYÁT OF OMAR KHAYYÁM, TRANSLATION BY EDWARD FITZGERALD

THE UNIVERSE IS POPULATED with brilliant stars illuminating exotic planets and fabulously beautiful clouds of glowing gas. But other than the objects in our tiny local Solar System, they are all out of reach for the foreseeable future. No human space probe has visited another star, and no telescope can directly examine the insides of any celestial object. The information obtained about most of the Universe is contained in the light reaching Earth across space.

Earthbound humans knew almost nothing about the composition of celestial objects until the early 19th century. First, the German optician Joseph von Fraunhofer studied the spectrum of the Sun and discovered that it is interrupted by more than 600 narrow dark lines that are colors missing from the sunlight that Earth receives. Then other scientists performed laboratory experiments showing that those spectral lines are related to the presence of various atoms in the Sun's atmosphere. Finally, astronomers observed that the spectra of other stars have similar patterns of lines, opening a window to real understanding of how the Sun and stars are related.

In this chapter, you will see how the Sun and other stars radiate light and how atoms in the atmospheres of stars, planets, and gas clouds in space interact with light to cause spectral lines (Figure 7-1). Once you understand that, you will know how astronomers determine the chemical composition of distant objects, as well as measure motions of gas in and around them.

7-1 Atoms

Atoms in stars and planets leave their fingerprints on the light we receive from them. By first reviewing what atoms are and then learning how they interact with light, you can understand how the spectra of objects in space are decoded.

7-1a A Model Atom

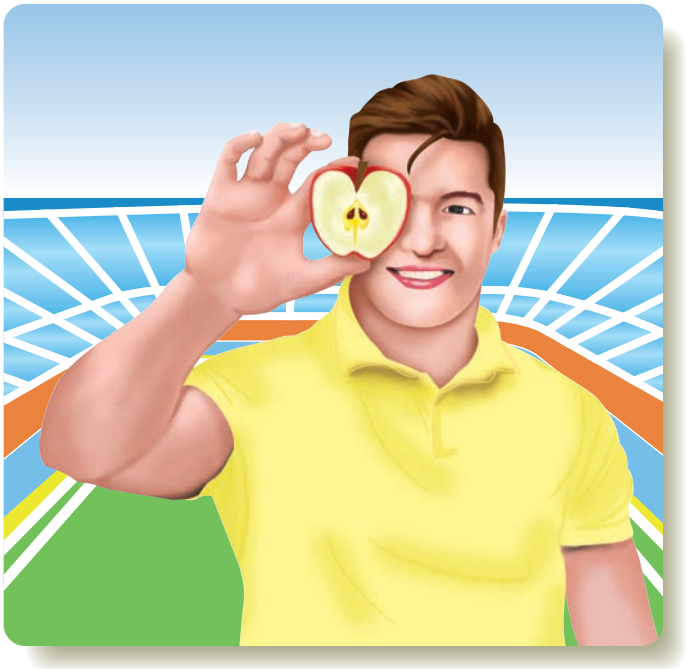
To think about how atoms interact with light, you need a working model of an atom. In Chapter 2, you used a model of the sky, the celestial sphere. In this chapter, you will begin your study of atoms by using a mental model of an atom. Remember that such a model can have practical value without being true. The stars are not actually attached to a sphere surrounding Earth, but to navigate a ship or point a telescope, it is convenient and practical to pretend they are. The electrons in an atom are not actually little beads orbiting the nucleus the way planets orbit the Sun, but for some purposes it is useful to picture them as such.

NASA/ESA/STScI/AURA/NSF

▲ **Figure 7-1** What's going on here? The sky is filled with beautiful and mysterious objects that lie far beyond your reach—in the case of the nebula NGC 2392, also called the Eskimo Nebula, about 5000 ly beyond your reach. The only way to understand such objects is by studying their light. Such analysis reveals that this object is a dying star surrounded by an expanding shell of gas it ejected a few thousand years ago.

A single atom is not a massive object. A hydrogen atom, for example, has a mass of only 1.7×10^{-27} kg, about a trillionth of a trillionth of a gram. Most of the mass of an atom lies in a **nucleus** that contains positively charged **protons** and uncharged **neutrons**, which have masses almost 2000 times greater than that of negatively charged **electrons**. Normally the number of electrons equals the number of protons so the positive and negative charges balance to produce a neutral atom. In this atomic model, the electrons can be pictured as being in a cloud that completely surrounds the nucleus.

An atom is mostly empty space. To see this, imagine constructing a simple scale model of a hydrogen atom. Its nucleus is a single proton with a diameter of approximately 0.0000017 nm, or 1.7×10^{-15} m. If you multiply that by 1 trillion (10^{12}), you can represent the nucleus of your model atom with something about 1.7 mm in diameter—an apple seed would do. The region of a hydrogen atom that normally contains the electron has an effective diameter of about 0.24 nm, or 2.4×10^{-10} m. This is two times a hydrogen atom's Van der Waals radius, which characterizes the distance over which it interacts with other atoms. Multiplying that by a trillion increases the diameter to about 240 m, or about the size of a football stadium (Figure 7-2). When you imagine an apple seed in the middle of a sphere the size of a football stadium, you can see that an atom is mostly empty space.



▲ **Figure 7-2** Magnifying a hydrogen atom by 10^{12} makes the nucleus the size of an apple seed and the electron cloud about the size of a stadium.

Now you can consider a **Common Misconception**. Most people, without thinking about it much, imagine that matter is solid, but you have seen that atoms are mostly empty space. The chair you sit on, the floor you walk on, are mostly not there. If you study the deaths of stars in a later chapter, you will see what happens to a star when the empty space gets squeezed out of its atoms.

7-1b Different Kinds of Atoms

There are more than a hundred chemical elements. The number of protons in the nucleus of an atom determines which element it is. For example, a carbon atom has six protons in its nucleus. An atom with one more proton than that is nitrogen, and an atom with one fewer proton is boron.

Although an atom of a given element always has the same number of protons in its nucleus, the number of neutrons is less restricted. For instance, if a neutron is added to a carbon nucleus, it would still be carbon, but it would be slightly heavier. Atoms that have the same number of protons but a different number of neutrons are **isotopes**. Carbon has two stable isotopes. One contains six protons and six neutrons for a total of 12 particles and is thus called carbon-12. Carbon-13 has six protons and seven neutrons in its nucleus.

The number of electrons in an atom of a given element can vary. Protons and neutrons are bound tightly into the nucleus, but the electrons are held loosely in the electron cloud. An atom that has lost or gained one or more electrons is called an **ion**. A neutral carbon atom has six electrons that balance the positive charge of the six protons in its nucleus. If you remove one or more electrons through **ionization**, the atom is left with a net positive charge. Under other circumstances, an atom may capture one or more extra electrons, giving it more negative charges than positive. Such a negatively charged atom is also considered an ion.

Atoms that collide may form bonds with each other by exchanging or sharing electrons. As you already know, two or more atoms bonded together form a **molecule**. Atoms do collide in stars, but the high temperatures cause violent collisions that are unfavorable for chemical bonding. Only in the coolest stars are the collisions gentle enough to permit the formation of chemical bonds. The presence of molecules such as titanium oxide (TiO) detected in some stars is one clue that those stars are very cool compared with other stars. In later chapters, you will see that molecules also can form in cool gas clouds in space and in the atmospheres of planets.

7-1c Electron Orbits

So far you have been considering the cloud of electrons in atoms only in a general way. Now it is necessary to be more specific about how electrons behave within the cloud as a means of understanding how light interacts with atoms.

Electrons are bound to the atom by the attraction between their negative charge and the positive charge on the nucleus. This attraction is known as the **Coulomb force**, after the physicist Charles-Augustin de Coulomb. To ionize an atom, you need a certain amount of energy to pull an electron completely away from the nucleus. This energy is the electron's **binding energy**, the energy that holds it to the atom.

The size of an electron's orbit is related to the energy that binds it to the atom. If an electron orbits close to the nucleus, it is tightly bound, and a large amount of energy is needed to pull it away. In other words, its binding energy is large. An electron orbiting farther from the nucleus is held more loosely, and less energy is needed to pull it away. That means it has small binding energy.

Nature permits atoms only certain amounts (quanta) of binding energy. The laws that describe how atoms behave are called the laws of **quantum mechanics** (**How Do We Know? 7-1**). Much of this discussion of atoms is based on the laws of quantum mechanics that were discovered by physicists early in the 20th century.

Because atoms can have only certain amounts of binding energy, electrons can have orbits only of certain sizes, called **permitted orbits**. These are like steps in a staircase: You can stand on the number one step or the number two step, but not on the number one and one-half step because there isn't one. The electron can occupy any permitted orbit, but there are no orbits in between.

The arrangement of permitted orbits depends primarily on the charge of the nucleus, which in turn depends on the number of protons. Consequently, each chemical element—each type of atom—has its own pattern of permitted orbits (**Figure 7-3**). Isotopes of the same elements have nearly the same pattern because they have the same number of protons in their nuclei but slightly different masses. Ionized atoms, with altered electrical charges, have orbital patterns that differ greatly from their un-ionized forms.

Quantum Mechanics

How can you understand nature if it depends on the atomic world you cannot see?

You can see objects such as stars, planets, aircraft carriers, and hummingbirds, but you can't see individual atoms. As scientists apply the principle of cause and effect, they study the natural effects they can see and work backward to find the causes. Invariably, that quest for causes in the physical world leads back to the invisible world of atoms.

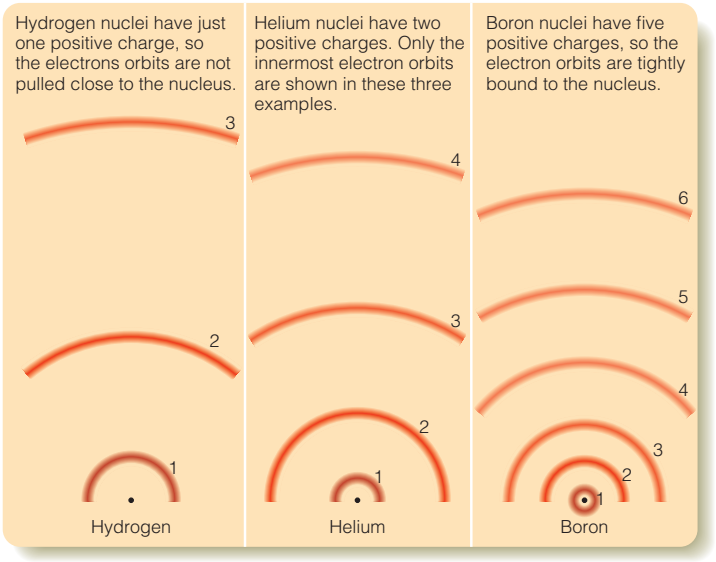
Quantum mechanics is the set of rules that describe how atoms and subatomic particles behave. On the atomic scale, particles behave in ways that seem unfamiliar and difficult to comprehend. One of the principles of quantum mechanics specifies that you cannot know simultaneously the exact location and exact motion of a particle. This is why, in practice, physicists go beyond the simple atomic model you may have learned in high school that imagines electrons as particles following orbits, and instead describe the electrons in an atom as if they are each clouds of negative charge surrounding the nucleus. That's a much better model, although it's still only a model.

This raises some serious questions about reality. Is an electron really a particle at all? If you can't know simultaneously the position and motion of a specific particle, how can you know how it will react to a collision with a photon or another particle? The surprising and puzzling answer is that you can't know certainly and completely. That seems to violate the principle of cause and effect.

Many of the phenomena you can see depend on the behavior of huge numbers of atoms, so quantum mechanical uncertainties average out. Nevertheless, the ultimate causes that scientists seek lie at the level of atoms, and modern physicists are trying to understand the nature of the particles that make up atoms. That is one of the most exciting frontiers of science.

Anna Henly/Getty Images

A display of aurora borealis, also known as "northern lights." Gas in Earth's upper atmosphere is excited by electrical currents caused by interactions of charged particles in the solar wind with Earth's magnetic field. This particular auroral glow shows spectral emission lines of ionized oxygen (*green*) and nitrogen (*red*).



▲ **Figure 7-3** An electron in an atom may occupy only certain permitted orbits. Because each element has a different number of protons and therefore a different electrical charge in the nucleus attracting the electrons, each element has a different, unique pattern of permitted orbits.

Practicing **SCIENCE**

How many hydrogen atoms would it take to cross the head of a pin? By answering this question, you will discover how small atoms really are and also see how important mathematics are in doing science and understanding nature.

To begin, assume that the head of a pin is about 1 mm in diameter—that is, 0.001 m. The size of a hydrogen atom is represented by the diameter of the electron cloud, roughly 0.24 nm. Because 1 nm equals 10^{-9} m, you can multiply and discover that 0.24 nm equals 2.4×10^{-10} m. To find out how many atoms would stretch 0.001 m, you can divide the diameter of the pinhead by the diameter of an atom. That is, divide 0.001 m by 2.4×10^{-10} m, and you get 4.2×10^{6}. That means it would take 4.2 million hydrogen atoms lined up side by side to cross the head of a pin.

Now you can see how tiny an atom is and also how powerful a bit of simple math can be. It reveals a view of nature beyond the capability of your eyes. Now do some more science with another bit of math. ***How many hydrogen atoms would you need to add up to the mass of a paper clip (1 g)?***

7-2 Interactions of Light and Matter

If light and matter did not interact, you would not be able to see these words. In fact, you would not exist because, among other problems, photosynthesis would be impossible, so there would be no grass, wheat, bread, beef, yogurt, or any other kind of food. The interaction of light and matter makes life possible, and it also makes it possible for you to understand the Universe.

You have already been considering a model hydrogen atom. Now you can use that model as you begin your study of light and matter by thinking about hydrogen. It is both simple and common: Roughly 90 percent of all atoms in the Universe are hydrogen.

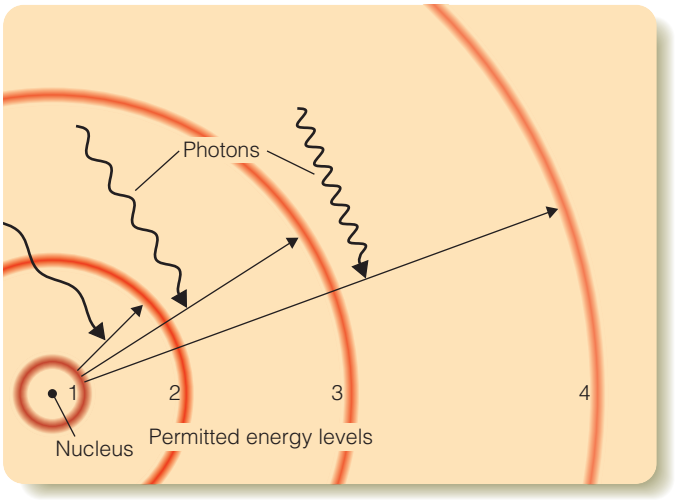
7-2a The Excitation of Atoms

Because each electron orbit in an atom represents a specific amount of binding energy, physicists commonly refer to the orbits as **energy levels**. Using this terminology, you can say that an electron in its smallest and most tightly bound orbit is in its lowest permitted energy level, which is called the atom's **ground state**. You could move the electron from one energy level to another by supplying enough energy to make up the difference between the two energy levels. It would be like moving a package from a low shelf to a high shelf; the greater the distance between the shelves, the more energy you would need to raise the package. The amount of energy needed to move the electron is the difference in the binding energy between the two levels. Giving the package a different amount of energy would put it between the shelves, where it would not be able to stay.

If you move an electron from a low energy level to a higher energy level, the atom becomes an **excited atom**. That is, you have added energy to the atom by moving its electron outward from the nucleus. One way an atom can become excited is by collision. If two atoms collide, one or both may have electrons knocked into a higher energy level. This happens very commonly in hot gas, where atoms move rapidly and collide often.

Another way an atom can become excited is to absorb a photon. As you learned in the previous chapter, a photon is a bundle of electromagnetic waves with a specific energy. Only a photon with exactly the right amount of energy can move the electron from one level to another. If the photon has too much or too little energy, that atom cannot absorb it. Because the energy of a photon depends on its wavelength, only photons of certain wavelengths (colors) can be absorbed by a given kind of atom.

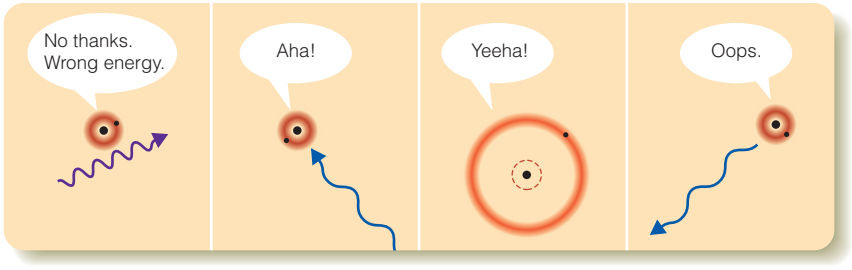
Figure 7-4 shows the lowest four energy levels of the hydrogen atom, along with three photons the atom is capable of absorbing. The photon with the longest wavelength has only enough energy to excite (move) the electron up to the second energy level, but the photons with shorter wavelengths have more energy and can excite the electron to higher levels. An actual hydrogen atom has many more energy levels than shown in Figure 7-4, and it can absorb photons of many different wavelengths.



▲ **Figure 7-4** A hydrogen atom can absorb only those photons that have the right energy to move the atom's electron to one of the higher-energy orbits. Here three photons with different wavelengths are shown along with the changes they each would produce in the electron's orbit if they were absorbed.

Atoms, like humans, cannot exist in an excited state forever. An excited atom is unstable and must eventually (usually within 10^{-9} to 10^{-6} seconds) give up the energy it has absorbed and return its electron to a lower energy level. Thus, the electron in an excited atom tends to tumble down to its lowest energy level, which is called its *ground state*. When an electron drops from a higher to a lower energy level, it moves from a loosely bound level to one that is more tightly bound. The atom then has a surplus of energy—the energy difference between the levels—that it can emit as a photon of light with a wavelength corresponding to that amount of energy (look back to Section 6-1a).

Study the sequence of events shown in **Figure 7-5** to see how an atom can absorb and emit photons. Here is the most important point in this chapter: Because each type of atom or ion has a unique set of energy levels, each type absorbs and emits photons with a unique set of wavelengths. As a result, you can



▲ **Figure 7-5** An atom can absorb a photon only if the photon has the correct amount of energy. The excited atom is unstable and within a fraction of a second returns to a lower energy level, radiating new photons in random directions relative to the original photon's direction.

identify the elements in a gas by studying the characteristic wavelengths of light that are absorbed or emitted.

The processes of atomic excitation and photon emission are common sights in urban areas at night. A neon sign glows when atoms of neon gas in a glass tube are excited by electricity flowing through the tube. As the electrons in the electric current flow through the gas, they collide with the neon atoms and excite them. Almost immediately after a neon atom is excited, its electron drops back to a lower energy level, emitting the surplus energy as a photon of a certain wavelength. The photons emitted by excited neon blend to produce a reddish-orange glow. Signs of other colors, generically called *neon signs,* contain other gases or mixtures of gases. (Look back to the image that opens this chapter.) Whenever you look at a "neon" sign, you are seeing atoms emitting energy in the form of photons with specific colors determined by the structure of electron orbits in those atoms.

Neon signs are simple, but stars are complex. Stars have colors, but those colors are not determined by the composition of the gases they contain. In the next section, you will discover why stars have different colors, and that will give you further insight into how light interacts with matter.

7-2b Radiation from a Heated Object

When you view the stars in the constellation Orion, you notice that they are not all the same color (look back to Figure 2-4). One of the Famous Stars, Betelgeuse, in the upper left corner of Orion, is quite red; another Famous Star, Rigel, in Orion's lower right corner, is blue. These differences in color arise from differences in temperature.

The starlight that you see comes from gases that make up the visible surface of the star, its photosphere. (Recall that you first learned about the photosphere of the Sun in Section 3-3c, in the context of solar eclipses.) Layers of gas deeper inside the star also emit light, but that light is reabsorbed before it can reach the surface. The gas above the photosphere is too thin to emit much light. The photosphere is the visible surface of a star because it is dense enough to emit lots of light but transparent enough to allow that light to escape.

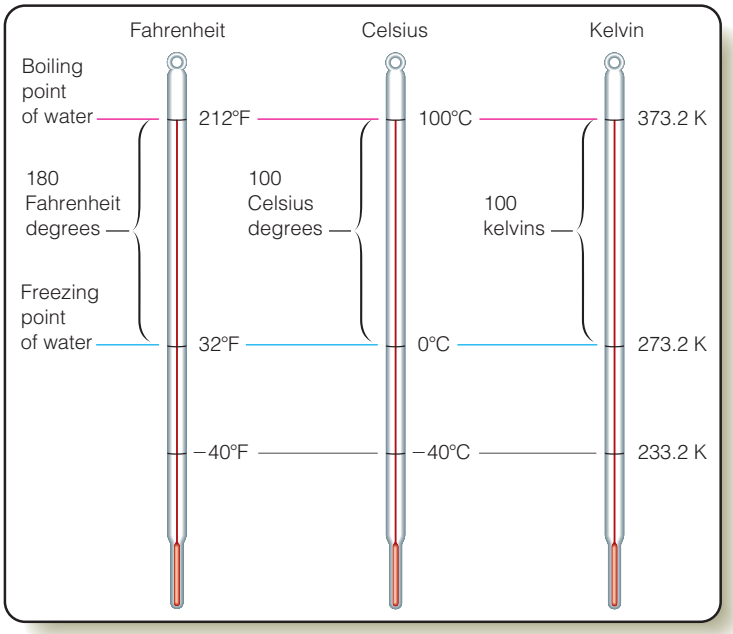
Stars produce their light for the same reason heated horseshoes glow in a blacksmith's forge—because they are hot. If a horseshoe is moderately hot, it glows ruddy red, but as it heats up further it grows brighter and yellower. Yellow hot is hotter than red hot, but not as hot as blue hot.

The light from stars and from glowing horseshoes is produced by the acceleration of charged particles. Usually the accelerated particles are electrons because they are the least massive charged particles, and they are on the outsides of atoms, so they are the easiest to get moving. An electron produces a surrounding electric field, and if an electron is accelerated, the change in its electric field spreads outward at the speed of light as electromagnetic radiation. You learned in Section 5-1a that *acceleration* means any change in motion—not only increasing speed, as that word means in everyday language, but also decreasing speed, and keeping constant speed while changing direction. Whenever the motion of any charged particle is changed, electromagnetic waves are generated. If you run a comb through your hair, you disturb electrons in both hair and comb, producing static electricity. That produces electromagnetic radiation, which you can hear as snaps and crackles if you are standing near an AM radio. Stars are hot, and they are made up of ionized gases, so there are plenty of electrons zipping around and being accelerated.

The molecules and atoms in any object are in constant motion, and in a hot object they are more agitated than in a cool object. This agitation is called **thermal energy**. It is a **Common Misconception** to think that *thermal energy, heat,* and *temperature* are the same. If you touch an object that contains lots of thermal energy, it will feel hot as the thermal energy flows into your fingers. The flow of thermal energy is called **heat**. In contrast, **temperature** refers to the average speed or intensity of agitation of the particles in an object. A metal speck from a fireworks sparkler can be much hotter than a hot metal clothes iron, but the iron contains much more thermal energy and therefore can burn your hand much more badly.

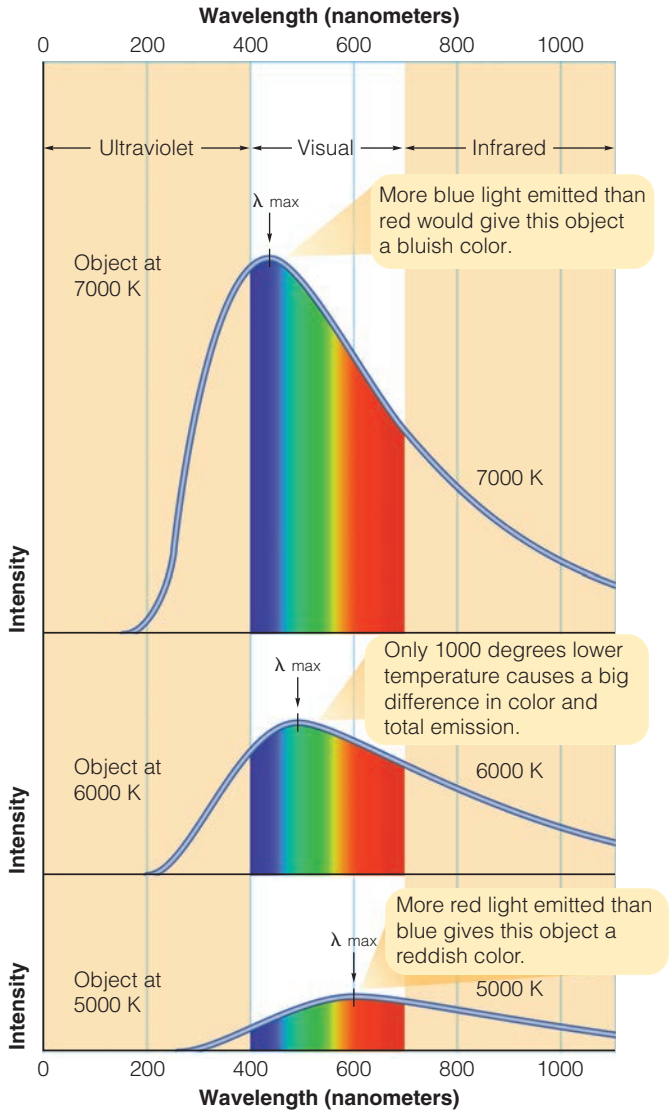
When astronomers refer to the temperature of a star, they are talking about the temperature of the gases in the photosphere, and they express those temperatures with the **Kelvin temperature scale** (Figure 7-6). On this scale, zero kelvins (written 0 K) is **absolute zero** (−273.2°C or −459.7°F), the temperature at which an object contains no thermal energy that can be extracted. Water at sea level atmospheric pressure freezes at 273.2 K and boils at 373.2 K. The Kelvin temperature scale is used in astronomy and physics because it is based on absolute zero and therefore is related most directly to the motions of particles in an object.

Now you can understand why a hot object glows or, to put it another way, why a hot object emits photons (bundles of



▲ **Figure 7-6** Comparison of three commonly used temperature scales.

electromagnetic energy). The hotter an object is, the more motion there is among its particles. The agitated particles, including electrons, collide with each other, and when electrons change their motion—accelerate—part of the energy of motion is carried away as electromagnetic radiation. The typical spectrum of an opaque heated object, meaning the amount and color distribution of radiation it emits, is called **blackbody radiation**. That name is translated from a German term referring to an object that is a perfectly efficient absorber and emitter of radiation. At room temperature, such a perfect absorber and emitter would look black, but at higher temperatures it would glow at wavelengths visible to a human eye. That explains why in later chapters you will see the term *blackbody* referring to objects that are actually glowing brightly.



▲ **Figure 7-7** Graphs of blackbody radiation intensity versus wavelength for three objects at temperatures of 7000 K, 6000 K, and 5000 K, respectively (*top to bottom*). Comparison of the graphs demonstrates that a hot object radiates more total energy per unit area than a cooler object (Stefan-Boltzmann law) and that the wavelength of maximum intensity is shorter for hotter objects than for cooler objects (Wien's law). The hotter object here would look blue to your eyes, whereas the cooler object would look red.

Blackbody radiation is quite common. In fact, it describes the light emitted by an incandescent lightbulb. Electricity flowing through the opaque filament of the bulb heats it to high temperature, and it glows with a blackbody spectrum. You can also recognize the light emitted by hot lava as blackbody radiation. Many objects in the sky, including the Sun and other stars, emit radiation approximately as blackbodies because they are mostly opaque.

Hot objects emit blackbody radiation, but so do cold objects. Ice cubes are cold, but their temperature is higher than absolute zero, so they contain some thermal energy and must emit blackbody radiation. The coldest gas drifting in space has a temperature only a few degrees above absolute zero, but it also emits a blackbody spectrum.

7-2c Two Blackbody Radiation Laws

Two features of blackbody radiation are important. First, the hotter an object is, the more radiation it emits. Hot objects emit more radiation because their agitated particles collide more often and more violently with each other. That's why a glowing coal from a fire emits more total energy than an ice cube of the same size.

Figure 7-7 shows the intensity of radiation versus wavelength for three objects of different temperatures. The total area under each curve is proportional to the total energy emitted. You can see that the hottest object emits more total energy than the two cooler objects. This rule is known as the **Stefan-Boltzmann law**, named after Josef Stefan and Ludwig Boltzmann, the physicists who discovered it.

The Stefan-Boltzmann law mathematically relates the temperature of a blackbody to the total radiated energy. Recall from Section 5-2d that energy is expressed in units of joules (symbolized by capital J; the energy gained by an apple falling from a table onto the floor is approximately 1 joule). The total radiation in units of joules per second given off by 1 square meter of the surface of an object equals a constant number called the Stefan-Boltzmann constant, represented by the Greek lowercase letter sigma, σ, multiplied by the temperature in kelvins raised to the fourth power:

KEY EQUATION 7-1 Stefan-Boltzmann Law

$$E = \sigma T^4 \ (\text{J/s/m}^2)$$

For the sake of completeness, note that the constant σ equals 5.67×10^{-8} J/s/m^2/K^4, units of joules per second per square meter (equivalent to watts per square meter) per kelvin to the fourth power. Also note, a joule per second is a watt.

Example: How much energy is emitted by 1 square meter of the Sun's surface, at a temperature of 5780 K? ***Solution:***

$$\begin{aligned} E &= 5.67 \times 10^{-8} \times (5780)^4 \\ &= 5.67 \times 10^{-8} \times 1.12 \times 10^{15} \\ &= 6.35 \times 10^7 \ \text{J/s/m}^2 \end{aligned}$$

The second feature of blackbody radiation is the relationship between the temperature of the object and the wavelengths of the

photons it emits. The wavelength of the photon emitted when electrons collide with other particles depends on the violence of the collision; a violent collision can produce a short-wavelength (high-energy) photon. The electrons in an object have a distribution of speeds; a few move very rapidly, and a few move very slowly, but most travel at intermediate speeds. Because electrons with speeds much greater or much less than the average speed are rare, extremely violent collisions and extremely gentle collisions don't occur very often. Consequently, blackbody radiation is composed of a distribution of photons in which very short-wavelength and very long-wavelength photons are rare; photons with intermediate wavelengths are most common.

Look again at Figure 7-7, showing the intensity of blackbody radiation emitted versus wavelength for three objects of different temperatures. The curves are high in the middle and low at either end because the objects emit radiation most intensely at intermediate wavelengths. The **wavelength of maximum intensity (λ_{max})** is the wavelength at which the object emits the most intense radiation. (Note that λ_{max} refers not to the maximum wavelength but to the wavelength of the maximum.) You can see by comparing the three curves in Figure 7-7 that the wavelength of maximum intensity depends on temperature: The hottest object has the shortest wavelength of maximum emitted intensity. In other words, the hotter object emits more blue light than red and thus looks blue, and the cooler object emits more red than blue and consequently looks red. This rule is known as **Wien's law**, named after Wilhelm Wien.

Wien's law quantitatively expresses the relation between a blackbody's temperature and the wavelength of maximum intensity (λ_{max}) of its emitted spectrum. Written for conventional intensity units, the law is:

KEY EQUATION 7-2 Wien's Law

$$\lambda_{max} = 2.90 \times 10^6/T$$

That is, the wavelength, in nanometers, of the radiation with maximum intensity emitted by a blackbody equals 2.9 million divided by the blackbody's temperature on the Kelvin scale.

Example: What is the wavelength of maximum (strongest) emission from the surface of the Sun, at a temperature of 5780 K?

Solution:

$$\lambda_{max} = 2.90 \times 10^6/5780$$
$$= 5.02 \times 10^2 \text{ nm (502 nanometers)}$$

You can see both Wien's law and the Stefan-Boltzmann law in operation if you look down into a toaster after you start the toast. First, you see a faint deep red glow that, as the coils get hotter, becomes brighter (Stefan-Boltzmann law) and more orange-yellow in color (Wien's law). It's important to note that objects too cool to glow at visible wavelengths still produce blackbody radiation. For example, the human body has a temperature of 310 K and emits blackbody radiation mostly in the infrared part of the spectrum. Infrared security cameras can detect intruders by the radiation they emit, and mosquitoes can track you down in total darkness by homing in on your infrared radiation. Although you emit lots of infrared radiation, you almost never emit a gamma-ray or radio photon. The wavelength of maximum intensity of your glow lies in the infrared part of the spectrum.

How do the Stefan-Boltzmann law and Wien's law help you understand stars and other celestial objects? Suppose a star the same size as the Sun has a surface temperature twice as hot as the Sun's surface. Then, according to the Stefan-Boltzmann law, each square meter of that star radiates $2^4 = 16$ times as much energy as a square meter of the Sun's surface. You can see that a small difference in temperature between two stars can produce a large difference in the amount of energy emitted from their surfaces. The Stefan-Boltzmann law relates temperature, surface area, and total energy emitted by stars and other blackbodies. If you know two of those three quantities, you can determine the other one.

Using Wien's law, you can measure the temperatures of distant objects without having to travel there and stick in a thermometer. A cool star with a temperature of 2900 K will emit most intensely at a wavelength of 1000 nm, which is infrared light. In comparison, a very hot star with a temperature of 29,000 K radiates most intensely at a wavelength of 100 nm, which is ultraviolet light. Now you can understand why the two Famous Stars in Orion mentioned previously, Betelgeuse and Rigel, have such different colors. Betelgeuse is relatively cool and therefore looks red, but Rigel is hot and looks blue.

Practicing **SCIENCE**

The infrared radiation coming out of your ear can tell a doctor your temperature. How does that work? You know that your body is opaque and therefore radiates as a blackbody, and you know two blackbody radiation laws—the Stefan-Boltzmann law and Wien's law—but you need to understand them to pick the more useful one for this application.

Doctors and nurses can use a handheld device to measure body temperature by observing the infrared radiation emerging from a patient's ear. You might guess that the device depends on the Stefan-Boltzmann law and therefore measures the intensity of the infrared radiation. It is true that a person with a fever will emit more energy than a healthy person. However, a healthy person with a large ear canal (having more surface area) would emit more blackbody radiation than a person with a small ear canal, even if they have the same temperature, so measuring intensity would not necessarily be helpful. The medical device actually depends on Wien's law, finding temperature by measuring the "color" of the infrared radiation. A patient with a fever will emit with a slightly shorter wavelength of maximum intensity, and the infrared radiation emerging from his or her ear will be a tiny bit "bluer" than that emitted by a person with a normal temperature.

7-3 Understanding Spectra

Science is a way of understanding nature, and the spectrum of a star can tell you a great deal about the star's temperature, motion, and composition. In later chapters, you will use spectra to study other astronomical objects such as galaxies and planets, but you can begin by looking at the spectra of stars, including that of the Sun.

The spectrum of a star is formed as light passes outward through the gases near its surface. Read **Atomic Spectra** and notice that it describes important properties of spectra and defines 10 new terms that will help you understand astronomical spectra:

1. There are three types of spectra: (i) *continuous spectra;* (ii) *emission spectra,* which contain *emission lines;* and (iii) *absorption spectra,* which contain *absorption lines.* These types of spectra are described by *Kirchhoff's laws.* When you see one of these types of spectra, you can recognize the arrangement of matter that emitted the light.
2. Photons are emitted or absorbed when an electron in an atom makes a *transition* from one energy level to another. The wavelengths of the photons depend on the energy difference between the two levels, so (note, this is especially important) each spectral line represents not one energy level but rather an electron transition between two energy levels. Hydrogen atoms can produce many spectral lines that are grouped in series such as the *Lyman, Balmer,* and *Paschen series*. Only three hydrogen lines—all in the Balmer series—are visible to human eyes. The emitted photons coming from a hot cloud of hydrogen gas have the same wavelengths as the photons absorbed by hydrogen atoms in a cool cloud between an observer and a light source.
3. Most modern astronomy books and articles display spectra as graphs of intensity versus wavelength. Be sure you recognize the connection between dark absorption lines, bright emission lines, and the dips and peaks in the graphed spectrum.

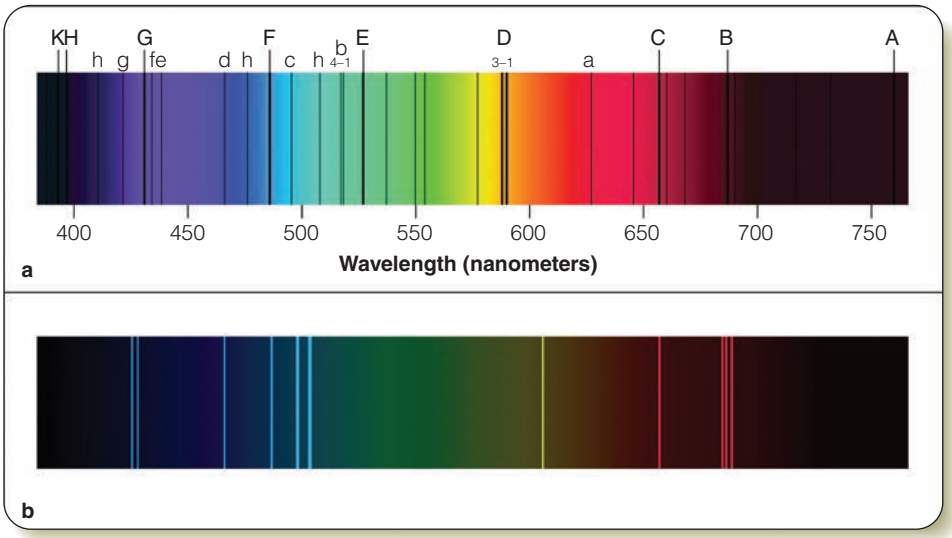
Imagine you are an astronaut with a handheld spectrograph, approaching a fresh lava flow on a planet with no atmosphere. You aim your spectrograph straight at the lava flow; what kind of spectrum do Kirchhoff's laws say you will see? You should see a continuous (blackbody) spectrum, with all the colors of the rainbow present, produced by the opaque, glowing hot lava. And as a bonus, measuring the wavelength of the strongest blackbody emission lets you determine the temperature of the lava using Wien's law.

Suddenly, the lava flow begins to bubble, and gas trapped in the molten rock is released, making a temporary, warm, thin atmosphere right above the lava flow. If you point your spectrograph to look through that gas with the hot lava as the background, you will observe an absorption (dark-line) spectrum in which the lava's blackbody spectrum is now interrupted by missing colors caused by atoms in the gas absorbing photons on their way from the lava to you. Finally, you crouch down and point your spectrograph at the gas at such an angle that the background is not hot lava but cold, empty, dark sky. Now you will see an emission (bright-line) spectrum, produced by atoms in the gas releasing photons as their electrons drop down toward the ground state, with lines at the same wavelengths as in the absorption spectrum of the same gas.

7-3a Chemical Composition

Identifying the elements that are present in a star, planet, or gas cloud by identifying the lines in that object's spectrum is a relatively straightforward procedure. For example, two dark absorption lines appear in the yellow region of the Sun's spectrum at the wavelengths 589.0 nm and 589.6 nm. The only atom that can produce this pair of lines is sodium, so the Sun must contain sodium. More than 90 elements in the Sun have been identified this way (**Figure 7-8a**).

However, just because the spectral lines that identify an element are missing, you cannot conclude that the element itself is absent. For example, the spectral lines in the hydrogen Balmer series are weak in the Sun's spectrum, even though 90 percent of



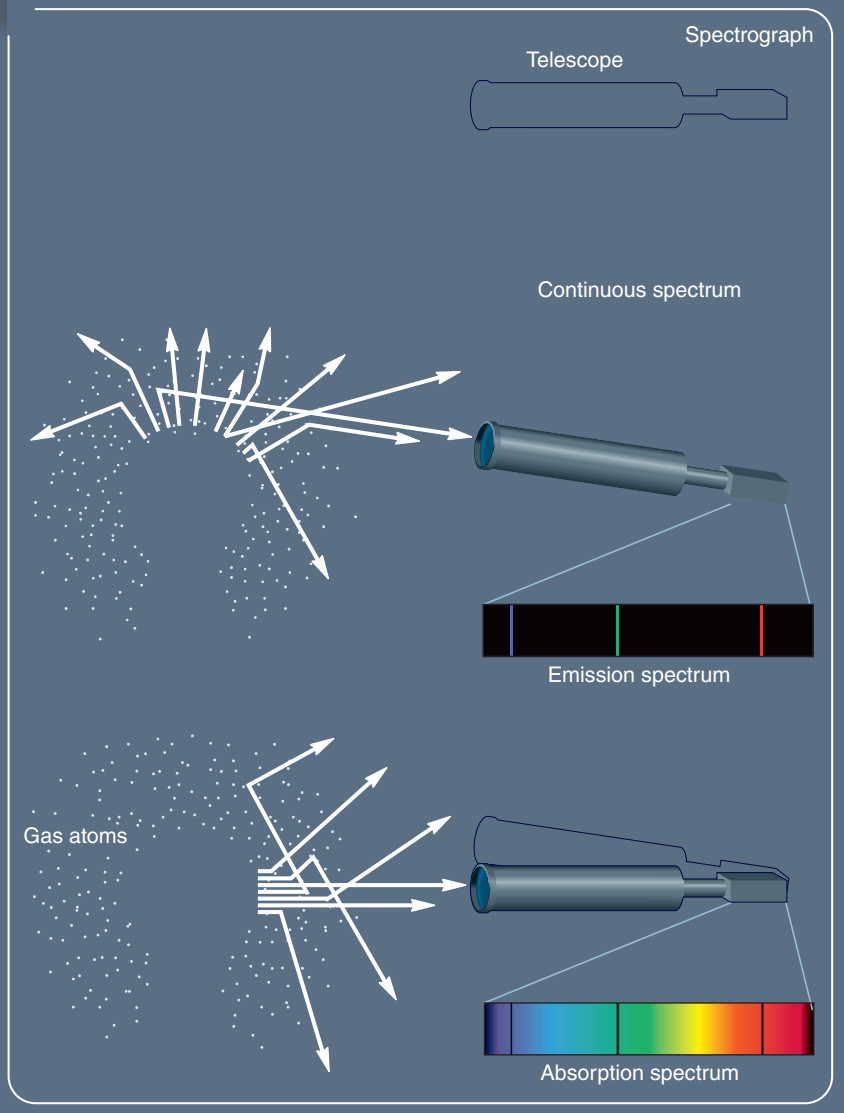
▲ **Figure 7-8** (a) The Sun's spectrum at visual wavelengths. The bright colored background shows the continuous spectrum of blackbody emission from the Sun's photosphere. The dark spectral absorption lines represent precise colors (photons of exact energies) removed from the Sun's radiation by atoms in its transparent atmosphere. (b) A model of the visual-wavelength emission (bright-line) spectrum of NGC 2392, the nebula shown in Figure 7-1. Emission lines from ionized atoms of hydrogen (*red*), nitrogen (*red*), and oxygen (*green*), among others, are seen here.

Atomic Spectra

1 To understand how to analyze a spectrum, begin with a simple incandescent lightbulb. The hot filament emits blackbody radiation, which forms a **continuous spectrum**.

An **emission spectrum** is produced by photons emitted by an excited gas. You could see **emission lines** by turning your telescope aside so that photons from the bright bulb do not enter the telescope and the excited gas has a dark background. The photons you would see would be those emitted by the excited atoms near the bulb, and the observed spectrum is mostly dark with a few bright emission lines. Such spectra are also called *bright-line spectra*.

An **absorption spectrum** results when radiation passes through a cool gas. In this case you can imagine that the lightbulb is surrounded by a cool cloud of gas. Atoms in the gas absorb photons of certain wavelengths that are then missing from the observed spectrum, and you see dark **absorption lines** at those wavelengths. Such absorption spectra are also called *dark-line spectra*.



1a The spectrum of a star is an absorption spectrum. The denser layers of the photosphere emit blackbody radiation. Gases in the atmosphere of the star absorb their specific wavelengths and form dark absorption lines in the spectrum.

Absorption spectrum

1b In 1859, long before scientists understood atoms and electron energy levels, the German scientist Gustav Kirchhoff formulated three rules—now known as **Kirchhoff's laws**—describing the three types of spectra.

KIRCHHOFF'S LAWS

Law I: The Continuous Spectrum

A solid, liquid, or dense gas excited to emit light will radiate at all wavelengths and thus produce a continuous spectrum.

Law II: The Emission Spectrum

A low-density gas excited to emit light will do so at specific wavelengths and thus produce an emission spectrum.

Law III: The Absorption Spectrum

If light comprising a continuous spectrum passes through a cool, low-density gas, the result will be an absorption spectrum.

Courtesy of Historica/Library of Congress/Diomedia

954.6 nm
1005.0 nm
1093.8 nm
1281.8 nm
1875.1 nm

388.9 nm
397.0 nm
410.2 nm
434.0 nm
486.1 nm
656.3 nm

93.8 nm
95.0 nm
97.3 nm
102.6 nm
121.6 nm

Paschen series (IR)

H_ζ

H_β
H_α

Balmer series (Visible-UV)

Lyman series (UV)

Nucleus

2 Electron orbits in the hydrogen atom are shown here as energy levels. When an electron makes a **transition** from one orbit to another, this means that the energy stored in the atom has changed. In this diagram, arrows pointed inward toward the nucleus represent transitions that result in the emission of a photon. If the arrows pointed outward, they would represent transitions that result from the absorption of a photon. Long arrows represent large amounts of energy and correspondingly short-wavelength photons.

2a Transitions in the hydrogen atom can be grouped into series—the **Lyman series, Balmer series, Paschen series**, and so on, named after scientists who carefully investigated the spectra of hydrogen atoms. Transitions and the resulting spectral lines are identified by Greek letters. Only the first few transitions in the first three series are shown at left.

2000 nm
1500 nm
Infrared
Paschen lines

2b In this drawing (*right*) of the hydrogen spectrum, emission lines in the infrared and ultraviolet are shown as gray. Only the first three lines of the Balmer series are visible to human eyes.

1000 nm

2c Excited clouds of gas in space emit light at all of the Balmer wavelengths, but you see only the red, blue, and violet photons blending to create the purple-pink color typical of ionized hydrogen.

The shorter-wavelength lines in each series blend together.

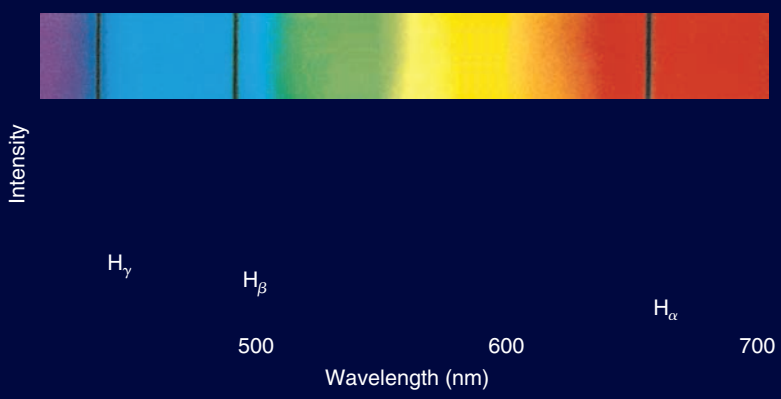
H_α

AURA/NOAO/NSF

Visual

500 nm
Visible
Balmer lines
H_β
H_γ

3 Modern astronomers rarely work with spectra in the form of images of bands of light. Spectra are usually recorded digitally, so it is easy to represent them as graphs of intensity versus wavelength. Here, the artwork above the graph suggests the appearance of a stellar spectrum. The graph at right reveals details not otherwise visible and allows comparison of relative intensities. Notice that dark absorption lines in the spectrum appear as dips in the intensity graph.



100 nm
Ultraviolet
Lyman lines

the atoms in the Sun are hydrogen. The next chapter will explain how it was discovered that this occurs because the Sun is too cool to produce strong hydrogen Balmer lines. Astronomers must consider that an element's spectral lines may be absent from an object's spectrum not because that element is missing but because that object has the wrong temperature to excite those atoms to the energy levels that produce detectable spectral lines. The same principle applies when interpreting the spectrum of a transparent gas cloud (Figure 7-8b). Seeing a particular emission line means the type of atom which produces that line is present in the cloud, but not seeing a line from some element's atoms does not necessarily mean the element is missing. The temperature of the cloud may be too hot or too cold to produce that particular spectral line.

To derive accurate chemical abundances, astronomers must use the laws that describe the interaction of light and matter to analyze a spectrum, take into account the object's temperature, and calculate the amounts of the elements present there. Such results show that nearly all stars, clouds of gas in interstellar space, and most of the visible matter in the Universe have a chemical composition similar to the Sun's—about 91 percent of the atoms are hydrogen, and 8.9 percent are helium, with small traces of heavier elements. You will use these results in later chapters when you study the life stories of the stars, the history of our Galaxy, and the origin of the Universe.

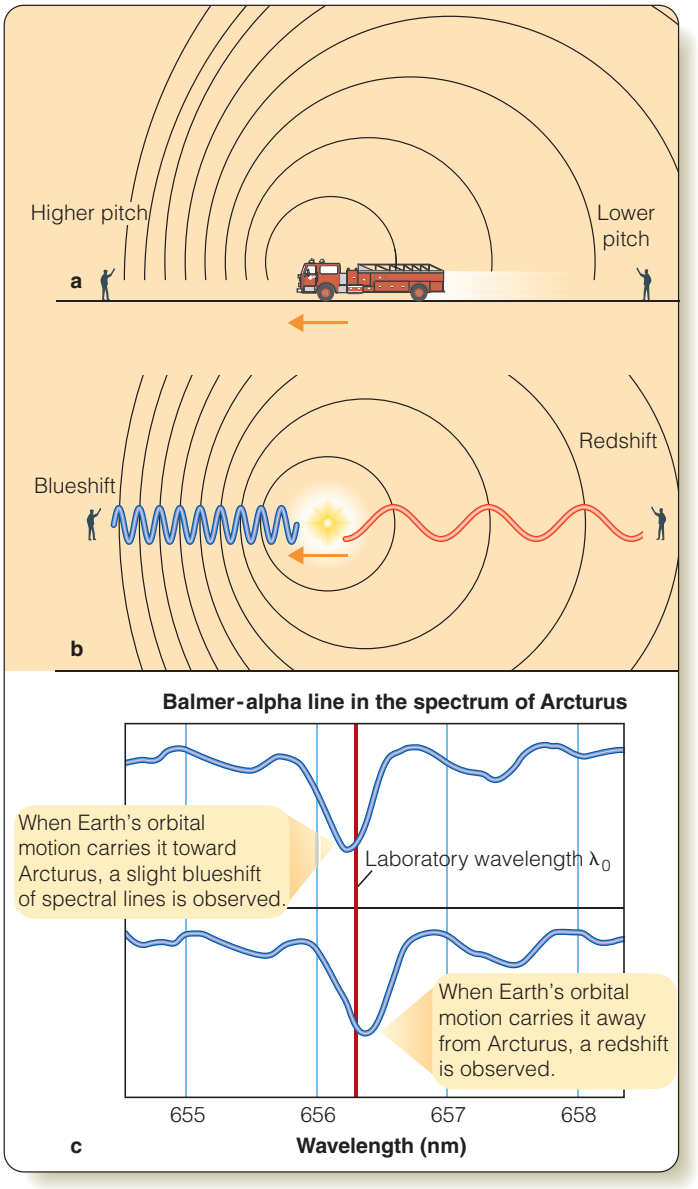
7-3b Measuring Velocities—The Doppler Effect

Surprisingly, one of the pieces of information hidden in a spectrum is the velocity of the light source. Astronomers can measure the wavelengths of the lines in a star's spectrum and find the velocity of the star. The **Doppler effect** is the apparent change in the wavelength of radiation from a source caused by the relative motion of the source and observer.

When astronomers talk about the Doppler effect, they are talking about a shift in the wavelength of electromagnetic radiation. But the Doppler shift can occur in any type of wave phenomena, including sound waves. You probably hear the Doppler effect several times every day without noticing. Every time a car or truck passes you and the pitch of its engine noise seems to drop, that's the Doppler effect. The pitch of a sound is determined by its wavelength; sounds with long wavelengths have low pitches, and sounds with short wavelengths have higher pitches. The vehicle's sound is shifted to shorter wavelengths and higher pitches while it is approaching to longer wavelengths and lower pitches after it passes as a result of the Doppler effect.

To see why the sound waves are shifted in wavelength, consider a fire truck approaching you with its siren blaring (Figure 7-9a). The sound coming from the siren will be a wave that can be depicted as a series of "peaks" and "valleys" representing compressions and decompressions. If the truck and siren are moving toward an observer, the peaks and valleys of the siren's sound wave will arrive closer together—at a higher frequency—than if the truck were not moving, and the observer will hear the siren at a higher pitch than the same siren when it is stationary. If the truck and siren are moving away, the sound waves will arrive farther apart—at a lower frequency, a lower pitch.

Now, substitute a source of light for the siren (Figure 7-9b). Imagine the light source emitting waves continuously as it approaches you. Each time the source emits the peak of a wave



▲ **Figure 7-9** The Doppler effect. (a) The sound waves (*black circles*) emitted from a siren on an approaching truck will be received more often, and thus be heard with a higher pitch, than the sound waves from a stationary truck. The siren will have a lower pitch if it is going away from the observer. (b) A moving source of light emits waves that move outward (*black circles*). An observer toward whom the light source is moving observes a shorter wavelength (a blueshift); an observer for whom the light source is moving away observes a longer wavelength (a redshift). (c) Absorption lines in the spectrum of the bright star Arcturus are blueshifted in winter, when Earth's orbital motion carries it toward the star, and redshifted in summer when Earth moves away from the star.

(meaning, the strongest electric and magnetic fields in the wave), it will be slightly closer to you than when it emitted the peak of the previous wave. From your vantage point, the successive peaks of the light wave will arrive closer together in the same way that the successive peaks of the siren's sound wave seemed closer together. The light will appear to have a shorter wavelength. Because shorter wavelengths are toward the blue, this is called a **blueshift**. After the light source has passed you and is moving away, the peaks of successive waves arrive farther apart, so the light has a longer wavelength and is redder. This is a **redshift**.

The terms *redshift* and *blueshift* are used to refer to any range of wavelengths. The light does not actually have to be red or blue visible light; the terms apply just as well to wavelengths in other parts of the electromagnetic spectrum such as X-rays and radio waves. *Red* and *blue* refer to the direction of the shift, not to actual color.

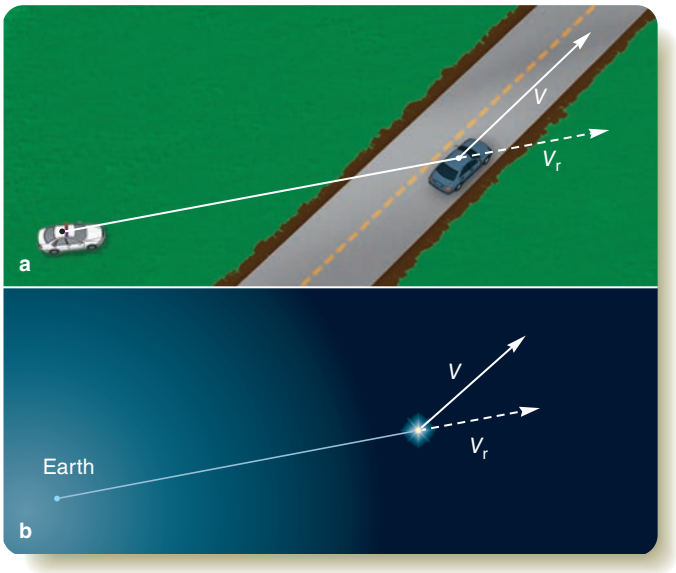
The amount of change in wavelength, and thus the size of the Doppler shift, depends on the velocity of the source. A moving car has a smaller Doppler shift than a jet plane, and a slow-moving star has a smaller Doppler shift than one that is moving more quickly. You can measure the velocity of a star by measuring the size of its Doppler shift. If a star is moving toward Earth, it has a blueshift, which means that each of its spectral lines is shifted toward shorter wavelengths. If it is receding from Earth, it has a redshift. The shifts are normally much too small to change the overall color of a star noticeably, but they are easily detected in spectra. In the next section, you will learn how astronomers can convert Doppler shifts into velocities.

When you think about the Doppler effect in relation to celestial objects, it is important to understand two things. Earth itself moves, so measurement of a Doppler shift really measures the relative motion between Earth and the object. Figure 7-9c shows the Doppler effect in two spectra of the star Arcturus. Lines in the top spectrum are slightly blueshifted because the spectrum was recorded when Earth, in the course of its orbit, was moving toward Arcturus. Lines in the bottom spectrum are redshifted because it was recorded six months later, when Earth was moving away from Arcturus. To find the true motion of Arcturus through space, astronomers must first account for the motion of Earth.

The second point to remember is that the Doppler shift is sensitive only to the part of the velocity directed away from you or toward you—the **radial velocity (V_r)**. You cannot use the Doppler effect to detect any part of the velocity that is perpendicular to your line of sight. That is why police using radar guns park right next to the highway (Figure 7-10a). They want to measure your full velocity as you drive toward them, not just part of your velocity. For the same reason, a star moving only across your field of view would have no blueshift or redshift because its distance from Earth would not be decreasing or increasing (Figure 7-10b).

7-3c Calculating Doppler Velocities

It is easy to calculate the radial velocity of an object from its Doppler shift. The formula is a simple proportion relating the



▲ **Figure 7-10** (a) Police radar can measure only the radial part of your velocity (V_r) as you drive down the highway, not your true velocity along the pavement (V). That is why police using radar should never park far from the highway. This police car is actually poorly placed to make a good measurement. (b) From Earth, astronomers can use the Doppler effect to measure the radial velocity (V_r) of a star, but that is less than its true total velocity, V, through space.

radial velocity V_r divided by the speed of light c, to the change in wavelength $\Delta\lambda$ of a line divided by the "laboratory" wavelength of the line, λ_0. (This simple version of the formula is used if the velocity is much less than the speed of light.) The laboratory wavelength λ_0 of a spectral line (subscript 0) is the wavelength it will have if the source of the light is not moving relative to the spectrograph. In the actual spectrum of a star, this spectral line's wavelength is observed shifted by some small amount, $\Delta\lambda$ (pronounced *delta-lambda;* delta conventionally symbolizes a small change). If the wavelength is increased (a redshift), $\Delta\lambda$ is positive; if the wavelength is decreased (a blueshift), $\Delta\lambda$ is negative. The radial velocity, V_r, of the star is given by the Doppler formula:

KEY EQUATION 7-3 Doppler Shift

$$\frac{V_r}{c} = \frac{\Delta\lambda}{\lambda_0}$$

Example: Suppose the laboratory wavelength λ_0 of a certain spectral line is 600.000 nm, but the line is observed in a star's spectrum at a wavelength $\lambda = 600.100$ nm. What is the star's radial velocity? ***Solution:*** First note that the change in wavelength $\Delta\lambda$ is +0.10 nm so that:

$$\frac{V_r}{c} = \frac{0.100}{600} = 0.000167$$

The radial velocity is 0.100/600 multiplied by the speed of light. In astronomy, velocities are almost always given in kilometers

per second, so the speed of light c is expressed in those units, as 3.00×10^5 km/s. Therefore, the radial velocity of this star is $(0.000167) \times 3.00 \times 10^5$ km/s), which equals 50 km/s. Because $\Delta\lambda$ is positive, you know the star is receding from you.

Armed with your new understanding of light and spectra, you are ready to focus on your first astrophysical object, the star that supports life on Earth—the Sun—which is the subject of the next chapter.

WHAT ARE WE?

Stargazers

Do you suppose chickens ever look at the sky and wonder what the stars are? Probably not. Chickens are very good at the chicken business, but they are not known for big brains and deep thought. Humans, in contrast, have highly evolved, sophisticated brains and are extremely curious. In fact, curiosity may be the most reliable characteristic of intelligence, and curiosity about the stars could have been the start of our ongoing attempts to understand the world around us.

For astronomers up to the time of Copernicus and Kepler, the stars were just points of light. There seemed to be no way to learn anything about them. Galileo's telescope revealed surprising details about the planets, but, even viewed through a large telescope, the stars are just points of light. Even when later astronomers began to realize that the stars were other suns, the stars seemed forever beyond human knowledge.

As you have seen, the key to understanding the Universe is knowledge about how light interacts with matter. In the past 150 years or so, scientists have discovered how atoms and light interact to produce the spectra we observe, and astronomers have applied those discoveries to the ultimate object of human curiosity—the stars.

Chickens may never wonder what the stars are, or even wonder what chickens are, but humans are curious animals, and we do wonder about the stars and about ourselves. Our yearning to understand the stars is just part of our quest to understand what we are.

Study and Review

SUMMARY

7-1 Atoms

- An atom consists of a **nucleus** surrounded by a cloud of **electrons**. The nucleus is made up of one or more positively charged **protons** and, except for hydrogen, uncharged **neutrons**.
- The number of protons in an atom determines which element it is. Atoms of the same element (that is, having the same number of protons) with different numbers of neutrons are called **isotopes**. A neutral atom is surrounded by a number of negatively charged electrons equal to the number of protons in the nucleus. An atom that has lost or gained an electron through **ionization** is called an **ion**. Two or more atoms joined together form a **molecule**.
- Electrons in an atom are attracted to the nucleus by the **Coulomb force**. As described by **quantum mechanics**, the **binding energy** that holds electrons in an atom is limited to certain energies, and thus electrons may occupy only certain **permitted orbits**.

7-2 Interactions of Light and Matter

- The size of an electron's orbit depends on its energy, so the orbits can be thought of as **energy levels**, with the lowest possible energy level known as the **ground state**. An **excited atom** is one in which an electron is raised to a higher orbit by a collision between atoms or the absorption of a photon having the proper energy.
- **Temperature** refers to the average intensity of the agitation among the atoms and molecules of an object that can be expressed on the **Kelvin temperature scale**, which gives temperature above **absolute zero**. The sum of the agitation of the particles in an object is called **thermal energy**, and the flow of thermal energy is **heat**.
- Collisions among the particles in a hot, dense object accelerate electrons and cause the emission of **blackbody radiation**. The **Stefan-Boltzmann law** states that the hotter an object, the more total energy the blackbody radiates:

$$E = \sigma T^4 (\text{J/s/m}^2) \qquad \text{(Eq. 7-1)}$$

Wien's law quantifies the relationship between a blackbody's temperature and the **wavelength of maximum intensity,** λ_{max}:

$$\lambda_{max} = 2.90 \times 10^6 / T \qquad \text{(Eq. 7-2)}$$

This principle allows astronomers to estimate the surface temperatures of stars from their colors.

7-3 Understanding Spectra

- **Kirchhoff's laws** summarize how (1) a hot solid, liquid, or dense gas emits electromagnetic radiation at all wavelengths and produces a **continuous spectrum**; (2) an excited, low-density gas produces an **emission spectrum** containing **emission lines**; and (3) a light source viewed through a low-density, cool gas produces an **absorption spectrum** containing **absorption lines**.
- An atom can emit or absorb a photon when an electron makes a **transition** between orbits.
- Because orbits of only certain energy differences are permitted in an atom, photons of only certain wavelengths can be absorbed or emitted. Each kind of atom has its own characteristic set of spectral lines. The hydrogen atom has the **Lyman series** of lines in the ultraviolet, the **Balmer series** partially in the visible, and the **Paschen series** (plus others) in the infrared.
- An emission or absorption spectrum can tell you the chemical composition of the stars. The presence of spectral lines of a certain element is evidence that element is present in the star, but the

strengths of the spectral lines of the chemical elements are not directly related to the elements' respective abundances. Lines of a certain element may be weak or absent in the observed spectra if the star is too hot or too cool, even if that element is present in the star's atmosphere.

- The **Doppler effect** can provide clues to the motions of the stars. When a star is approaching, you observe a chemical element with slightly shorter wavelengths than the pattern produced by that same chemical element in a lab, a **blueshift**. When a star is receding, you observe a chemical element with slightly longer wavelengths than the pattern produced by that same chemical element in the lab, a **redshift**. This Doppler effect reveals a star's **radial velocity, V_r**, the part of its velocity directed toward or away from Earth:

$$\frac{V_r}{c} = \frac{\Delta\lambda}{\lambda_0} \qquad \text{(Eq. 7-3)}$$

REVIEW QUESTIONS

1. Why might you say that an atom is mostly composed of empty space?
2. How many protons, neutrons, and electrons are in a neutral hydrogen atom? In a neutral helium atom? How many times heavier is the He atom compared to the H atom?
3. How is an isotope different from an ion?
4. Deuterium has one proton and one neutron in its nucleus, surrounded by one electron. Is deuterium an element, an ion, an isotope, and/or a molecule? Why? (More than one answer may be correct.)
5. A He-3 (helium-3) nucleus contains two protons and one neutron. If the atom is neutral, how many electrons surround the nucleus? Is He-3 an element, an ion, an isotope, and/or a molecule? Why? (More than one answer may be correct.)
6. Name a molecule. What atoms make up that molecule?
7. Why is the binding energy of an electron related to the size of the electron's orbit?
8. Explain why ionized calcium can form absorption lines, but ionized hydrogen cannot.
9. Describe two ways an atom can become excited.
10. An electron in a boron atom makes a transition from the fifth excited state to the third excited state. Did the atom become ionized as a result? Did the atom become excited as a result? Was a blackbody spectrum produced? How do you know?
11. Why do different atoms have different lines in their spectra?
12. Why does the amount of blackbody radiation emitted depend on the temperature of the object?
13. What is the wavelength of maximum intensity and the total energy emitted by a celestial object at absolute zero?
14. Why do hot stars look bluer than cool stars?
15. Why does a fireplace poker appear black when not in the fire and red, yellow, or white when left in the fire?
16. Celestial object A has a temperature of 60 K, and celestial object B has a temperature of 600 K. Which object emits the shorter wavelength of maximum intensity? Which objects has the least total energy emitted?
17. How is heat different from temperature?
18. What kind of spectrum does a neon sign produce? Describe this process. What colors are associated with a neon sign?
19. How can the Doppler effect explain wavelength shifts in both light and sound?
20. The emission spectra you obtained from a star shows a hydrogen spectrum that is shifted toward the blue end of the electromagnetic spectrum compared to the hydrogen spectrum you obtained from the lab. Is the star moving toward Earth, away from Earth, or is not enough information provided to determine its motion?
21. Which kind of spectrum is produced by a white household incandescent lightbulb?
22. Why does the Doppler effect detect only radial velocity?
23. Could an object be orbiting another object and we only detect the radial motion via the Doppler effect?
24. If the Doppler effect of light is a shift of spectral lines toward the blue or red end of the electromagnetic spectrum, what is the Doppler effect of sound?
25. **How Do We Know?** How is the macroscopic world you see around you determined by a microscopic world you cannot see?

ACTIVE INQUIRY QUESTIONS

1. Imagine that you work in a police lab and use a spectrograph to determine the chemical composition of samples from criminal investigations. Today, unfortunately, the power is out and none of your equipment works. Of the five basic human senses, which is the most useful in determining what a new incoming sample is made of?
2. Before Fraunhofer and others worked to observe and interpret spectra, most people were of the opinion that we would never know the composition of the Sun and stars. Can you think of any scientific question today that most people believe will probably never be answered?

PROBLEMS

1. Human body temperature is about 310 K (3.10×10^2 K). At what wavelength do humans radiate the most energy? In which part of the electromagnetic spectrum (gamma-ray, X-ray, UV, visible light, IR, microwave, or radio) do we emit?
2. Answer these questions for celestial bodies at each of the following temperatures and then draw a conclusion about the relationship between temperature and wavelength of maximum intensity. What is the wavelength of maximum intensity? In which part of the electromagnetic spectrum (gamma-ray, X-ray, UV, visible light, IR, microwave, or radio) does this peak wavelength lie? Give an example of an object that might have this temperature.
   a. 50 K
   b. 500 K
   c. 5000 K
   d. 50,000 K
3. If a star has a surface temperature of 20,000 K (2.00×10^4 K), at what wavelength will it radiate the most energy? Is this a cool or hot star?
4. Infrared observations of a star show that the star is most intense at a wavelength of 2000 nm (2.00×10^3 nm). What is the temperature of the star's surface?
5. If you double the temperature of a blackbody, by what factor will the total energy radiated per second per square meter increase?
6. If one star has a temperature of 6000 K and another star has a temperature of 7000 K, how much more energy per second will the hotter star radiate from each square meter of its surface?
7. What is the wavelength of maximum intensity and the total energy emitted by a celestial object at 5 K above absolute zero? Which part of the electromagnetic spectrum does the wavelength of maximum intensity lie?
8. Electron orbital transition A produces light with a wavelength of 500 nm. Transition B involves twice the energy of transition A. What wavelength is the light it produces?
9. Photon energy is related to wavelength by the Planck equation, $E = hc/\lambda$, where h is Planck's constant, 6.63×10^{-34} J s, and c is the

speed of light, 3.00×10^8 m/s. If the energy released by an electron making a transition from one hydrogen atom orbit to another is 1.64×10^{-18} J, what is the wavelength of the photon? Which part of the electromagnetic spectrum is that in? In which orbit did the electron start, and in which orbit did it finish? (*Hint:* Examine the Concept Art: Atomic Spectra.)

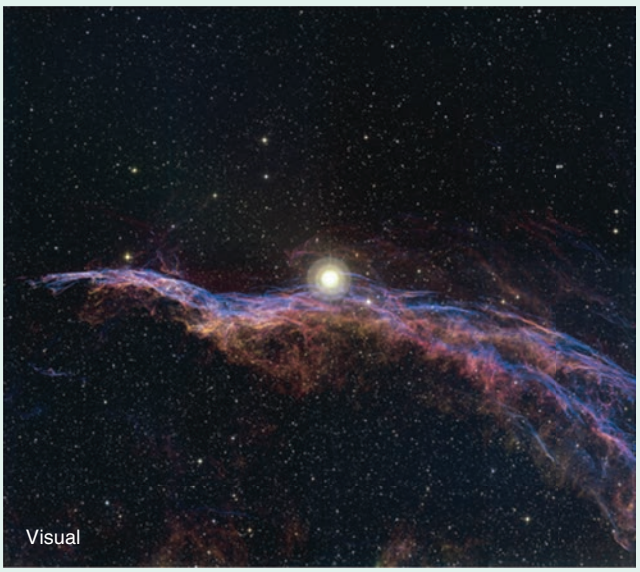
10. In a laboratory, the Balmer-beta spectral line of hydrogen has a wavelength of 486.1 nm. If the line appears in a star's spectrum at 486.3 nm, what is the star's radial velocity? Is it approaching or receding? Is this a blueshift or a redshift?
11. An astronomer observes the Balmer-beta line in a celestial object's spectrum at a wavelength of 496.5 nm. Is the object approaching or receding? If you can find the object's radial velocity, what is it? (*Note:* The laboratory wavelength of Balmer-beta is given in Problem 10.)
12. The highest-velocity stars an astronomer might observe in the Milky Way Galaxy have radial velocities of about 400 km/s (4.00×10^2 km/s). What change in wavelength would this cause in the Balmer-beta line? (*Note:* The laboratory wavelength of Balmer-beta is given in Problem 10.)

SENSE OF PROPORTION

1. The following objects all emit blackbody radiation. Arrange them in order of increasing wavelength of strongest emission: a hot iron, the Sun, your body, ice cubes
2. Reread the Practicing Science box at the end of Section 7-1. If a hydrogen atom with an actual diameter of 0.24 nm is represented by a baseball that is 2.9 in. in diameter, how big would the representation of a pinhead be? (Convert your answer to miles.)

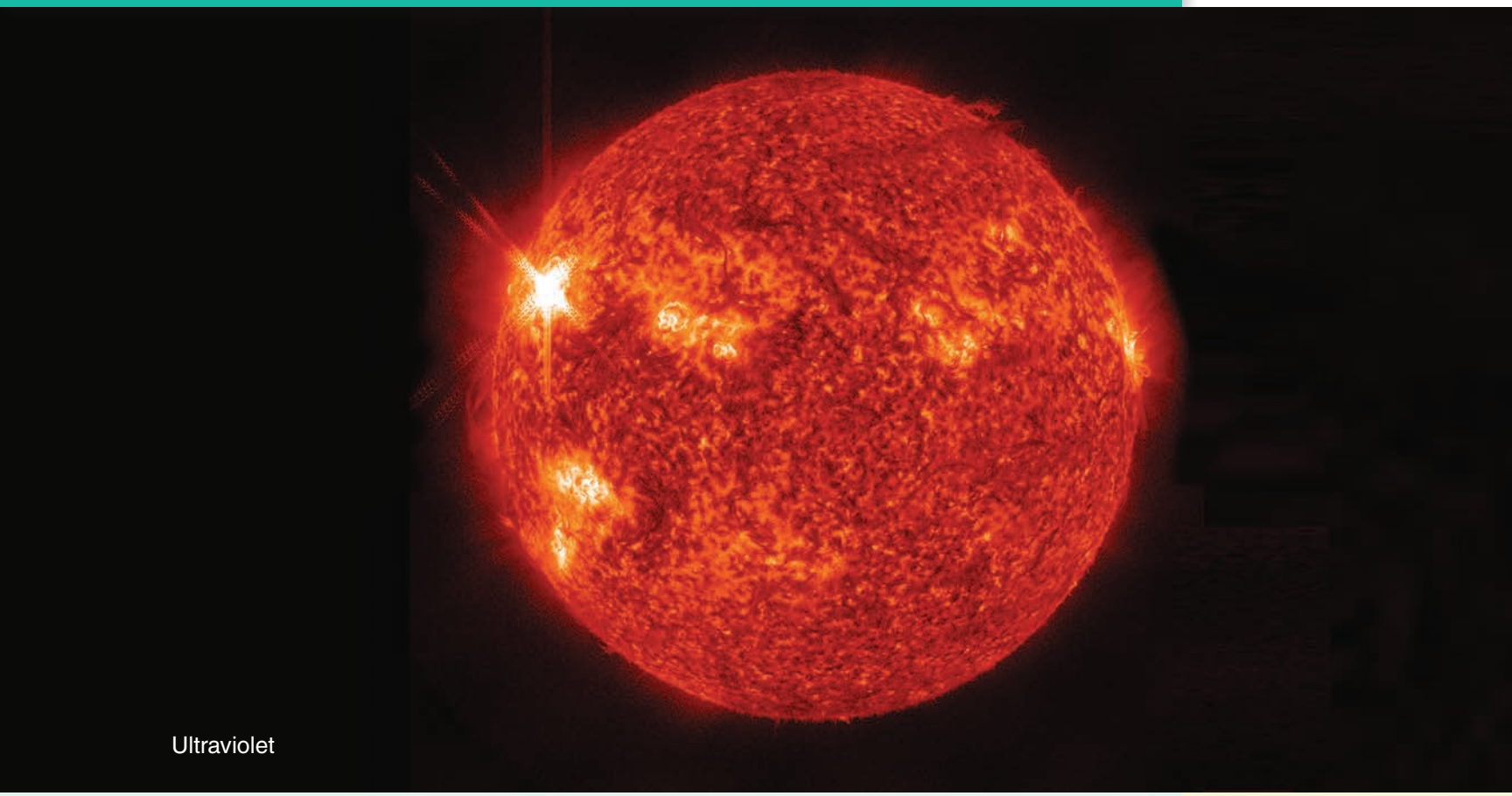
LEARNING TO LOOK

1. Consider Figure 7-3. Does an electron in the fifth excited state of a boron atom have more or less binding energy than the third excited state of helium? What about the electron in the first excited state of hydrogen compared to the second excited state of helium? Why?
2. Consider Figure 7-3. When an electron in a hydrogen atom moves from the third orbit to the second orbit, the atom emits a Balmer-alpha photon in the red part of the spectrum. In what part of the spectrum would you look to find the photon emitted when an electron in a helium atom makes the same transition?
3. Did ionized hydrogen produce the spectrum shown in Figure 3 of the Concept Art: Atomic Spectra?
4. What colors are the 589.0 nm and 589.6 nm sodium lines in the Sun's absorption spectrum, Figure 7-8a? Do these sodium lines shown in Figure 7-8a originate in the Sun's photosphere? How do you know?
5. Where should the police car in Figure 7-10a have parked to make a good measurement?
6. The nebula shown here contains mostly hydrogen excited to emit photons. What kind of spectrum would you expect this nebula to produce?
7. If the nebula shown here crosses in front of the star, and the nebula and star have different radial velocities, what might the spectrum of the star look like?



WIYN/NOAO/AURA/NSF/T.Rector(University of Alaska)

8 The Sun



NASA/*SDO*

▲ Ultraviolet image of the solar disk including a powerful flare (upper left) captured by the *Solar Dynamics Observatory* spacecraft. This flare was powerful enough to interfere with radio communications on Earth.

The Sun is the source of light and warmth in our Solar System, so it has always been a primary object of human curiosity and awe. It is also the star that is most easily visible from Earth. Understanding the interactions of light and matter that you studied in Chapter 7 can help reveal the secrets of the Sun and introduce you to the stars.

In this chapter, you will discover how analysis of the solar spectrum paints a detailed picture of the Sun's atmosphere and how basic physics has solved the mystery of what goes on in the Sun's core. Here you will find answers to four important questions:

- What can be learned about the Sun by observing its surface and atmosphere?
- What are the dark sunspots?
- Why does the Sun go through 11- and 22-year cycles of activity?
- What is the source of the Sun's energy?

Although this chapter considers only the star at the center of our Solar System, introducing you first to one star in detail lets you continue onward and outward in later chapters among the other stars that fill the Universe.

> **All cannot live on the piazza, but everyone may enjoy the sun.**
>
> ITALIAN PROVERB

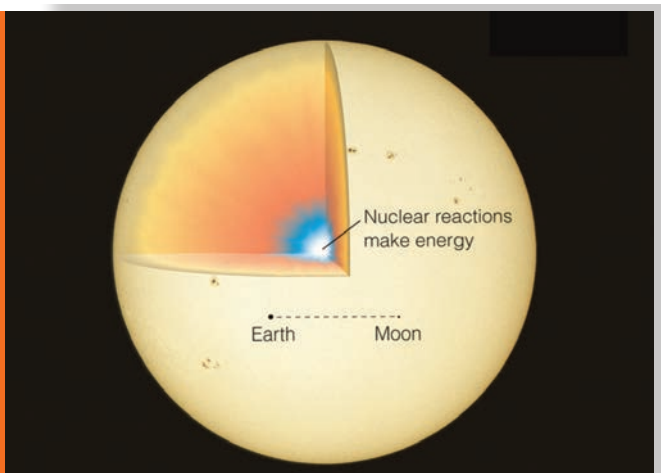
THE SUN IS so close compared with other stars that we can see swirling currents of gas and arched bridges of magnetic force with a level of detail that seems overwhelming. Solar astronomers sometimes remark ironically that they would know a lot more about the Sun if it were farther away. But, as you will learn in the next chapter, the Sun is just a normal star. In a sense, it is a simple object. The Sun is made up almost entirely of hydrogen and helium gas confined by its own gravity in a sphere 109 times Earth's diameter (**Celestial Profile: The Sun**). To give you a better sense of proportion about the size of the Sun, we'll compare features of the Sun to more familiar sizes on Earth throughout the chapter. The gases of the Sun's photosphere (surface) and atmosphere are hot, radiating the light and heat that are visible and make life possible on Earth. That part of the Sun is where you can begin your exploration.

8-1 The Solar Photosphere and Atmosphere

The part of the Sun you can see directly from Earth is made up of three layers. The visible surface is the **photosphere**, and above that are the **chromosphere** and the **corona**. (Note that astronomers normally speak of the interior of the Sun as being "below" or "under" the photosphere, and the Sun's atmosphere as being "above" or "over" the photosphere.) You first learned about these components of the Sun in the context of observing them during solar eclipses (look back to Section 3-3c).

When you view the Sun, you see the photosphere as a hot, glowing surface with a temperature of 5780 K. That temperature is determined by precisely measuring the spectrum of sunlight, then making a calculation using Wien's law (Eq. 7-2). At that temperature, every square millimeter of the Sun's surface is radiating more energy than a 60-watt lightbulb (Stefan-Boltzmann law, Eq. 7-1). With all that energy radiating into space, the Sun's surface would cool rapidly if new energy did not arrive to keep the surface hot, so simple logic tells you that there must be heat flowing outward from the Sun's interior. Not until the 1930s did astronomers understand that the Sun creates energy by nuclear reactions at its center. Those nuclear reactions are described in detail at the end of this chapter.

For now, you can consider the Sun's atmosphere in its relatively quiet, average state. Later you can add details regarding the types of activity produced by heat flow that makes the Sun's outer layers churn like a pot of boiling water.



Visual image + artist's conception

This image of the Sun shows a few sunspots and is cut away to show a diagram of the location of energy generation at the Sun's center. Earth, the Moon, and the distance between them are shown for scale.

Celestial Profile THE SUN

FROM EARTH

Average distance from Earth	1.000 AU (1.496×10^8 km)
Maximum distance from Earth	1.017 AU (1.521×10^8 km)
Minimum distance from Earth	0.983 AU (1.471×10^8 km)
Average apparent diameter	0.533° (1920 arc seconds)
Period of rotation (sidereal)	24.5 days at equator
Apparent visual magnitude	−26.74

PHYSICAL CHARACTERISTICS

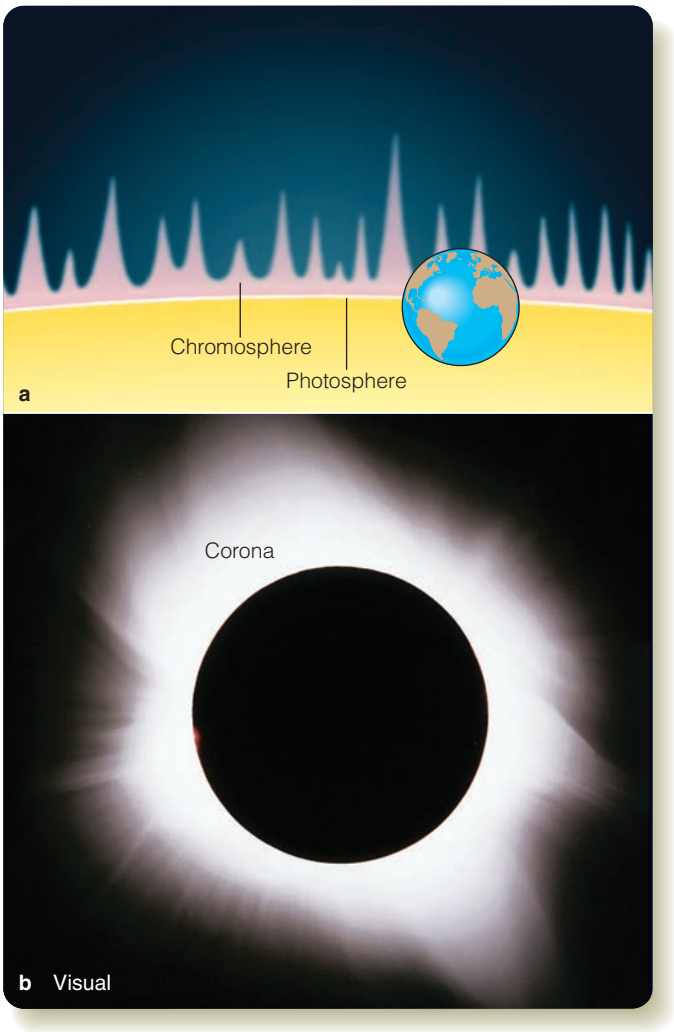
Radius	6.96×10^5 km
Mass	1.99×10^{30} kg
Average density	1.41 g/cm^3
Escape velocity at surface	618 km/s
Luminosity (electromagnetic)	3.84×10^{26} J/s
Surface temperature	5780 K
Central temperature	15.7×10^6 K

PERSONALITY PROFILE

In Greek mythology, the Sun was carried across the sky in a golden chariot pulled by powerful horses and guided by the Sun god, Helios. When Phaeton, the son of Helios, drove the chariot one day, he lost control of the horses, and Earth was nearly set ablaze before Zeus smote Phaeton from the sky. Even in classical times, people understood that life on Earth depends critically on the Sun.

8-1a The Photosphere

The visible surface of the Sun seems to be a distinct boundary, but it is not solid. In fact, the Sun is gaseous from its outer



▲ **Figure 8-1** (a) A cross section at the edge of the Sun shows the relative thickness of the photosphere and chromosphere. Earth is shown for scale. On this scale, the disk of the Sun would be more than 1.5 m (5 ft) in diameter. The corona extends from the top of the chromosphere to a great distance above the photosphere. (b) This photograph, made during a total solar eclipse, shows only the inner part of the corona.

atmosphere right down to its center. The photosphere is a thin layer of gas from which Earth receives most of the Sun's light (**Figure 8-1**). It is less than 500 km (300 mi) deep. In a model of the Sun the size of a bowling ball, the photosphere would be no thicker than a layer of tissue paper wrapped around the ball. (On that scale, Earth would be $^{1}/_{12}$ of an inch in diameter, about the size of an apple seed.)

The photosphere is the layer in the Sun's atmosphere that is dense enough to emit plenty of light but not so dense that light can't escape. Below the photosphere, the gas is denser and hotter and therefore radiates plenty of light, but that light cannot escape because it is blocked by the outer layers of gas. Therefore, Earth does not receive light from deeper layers under the photosphere. In contrast, the gas above the photosphere is less dense, so although Earth does receive that light, there is not much of it.

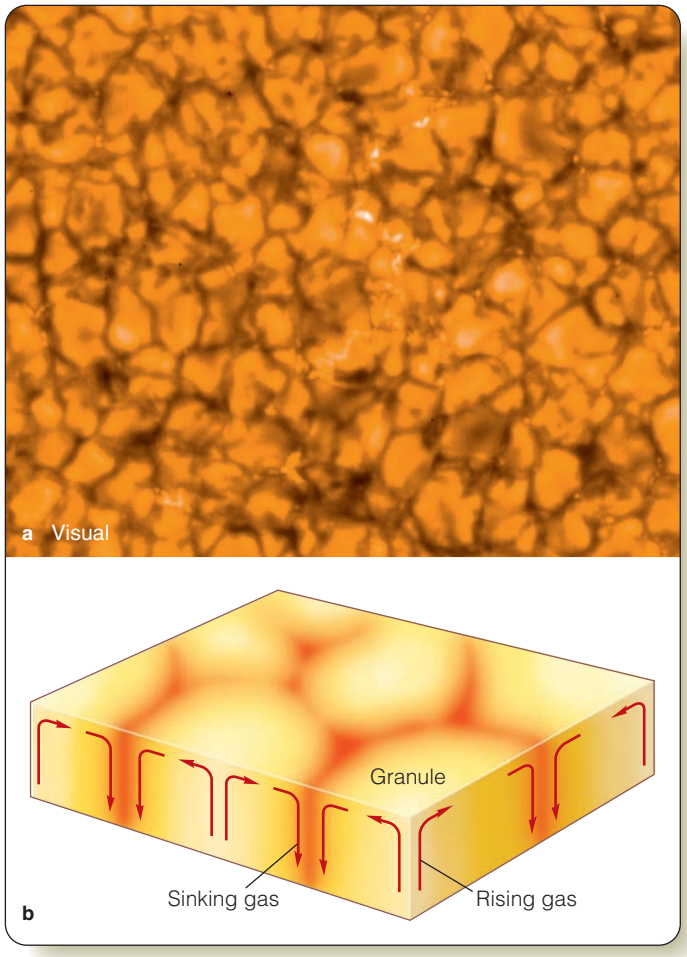
The photosphere appears to be substantial, but it is really a very low-density gas. Even in its deepest and densest layers, the photosphere is less than 1/3000 as dense as the air you breathe. To find gases as dense as the air at Earth's surface, you would have to descend about 15,000 km (10,000 mi) below the photosphere. With fantastically efficient insulation, you could fly a spaceship right through the photosphere.

The spectrum of the Sun is an absorption spectrum, and that can tell you a great deal about the photosphere. You know from Kirchhoff's third law (Chapter 7 Concept Art: Atomic Spectra) that an absorption spectrum is produced when the source of a continuous (blackbody) spectrum is viewed through a transparent gas. The deeper layers of the photosphere are dense enough to produce a continuous spectrum. Atoms in higher, transparent layers of the photosphere and in the Sun's atmosphere absorb photons with unique energies (wavelengths) corresponding to the jumps between each type of atom's electron orbits, producing absorption lines that allow you to identify hydrogen, helium, and other elements.

In high-resolution photographs, the photosphere has a mottled appearance because it is made up of dark-edged regions called *granules*. The overall pattern is called **granulation** (**Figure 8-2a**). Granules can be several thousand kilometers across, larger than the state of Texas, but last for only 10 to 20 minutes each before fading, shrinking, and being replaced by new granules. Doppler shifts of spectral lines reveal that the granule centers are rising and the edges are sinking at speeds of about 0.4 km/s (900 mph). Detailed observations of the granules show that their centers emit more blackbody radiation and are slightly bluer than the edges. Using the Stefan-Boltzmann and Wien laws (Eq. 7-1 and 7-2), astronomers can calculate that the granule centers are a few hundred degrees hotter than the edges (**Figure 8-2b**).

From this evidence, astronomers recognize granulation as the surface effects of **convection** currents just below the photosphere. Convection occurs when hot material rises and cool material sinks, as when, for example, a current of hot gas rises above a candle flame. You can watch convection in a liquid by adding a bit of cool creamer to an unstirred cup of hot coffee. The cool creamer sinks, gets warmer, expands, rises, cools, contracts, sinks again, and so on, creating small regions on the surface of the coffee that mark the tops of convection currents. Viewed from above, these coffee regions look something like solar granules. The presence of granulation is clear evidence that energy is flowing upward through the photosphere. You will learn more about the Sun's convection currents and internal structure later in this chapter.

Spectroscopic studies of the solar surface have revealed another larger but less obvious kind of granulation. **Supergranules** are regions a little over twice Earth's diameter that include an average of about 300 granules each. These supergranules are regions of very slowly rising currents that last a day or two. They appear to be produced by larger gas currents that begin deeper under the photosphere than the ones that produce the granules.



JAXA/NASA/PPARC/Hinode

▲ **Figure 8-2** (a) This ultra-high-resolution image of the photosphere shows granulation. The largest granules here are about 1500 km (1000 mi) across. (b) This model explains granulation as the tops of rising convection currents just below the photosphere. Heat flows upward as rising currents of hot gas and downward as sinking currents of cool gas. The rising currents heat the solar surface in small regions seen from Earth as granules.

The edge, or **limb**, of the solar disk is dimmer than the center (see the figure in Celestial Profile: The Sun). This **limb darkening** is caused by the absorption of light in the photosphere. When you look at the center of the solar disk, you are looking directly down into the Sun, and you see deeper, hotter, brighter layers in the photosphere. In contrast, when you look near the limb of the solar disk, you are looking at a steep angle to the surface and cannot see as deeply. The photons you see come from shallower, cooler, dimmer layers in the photosphere. Limb darkening proves that the temperature in the photosphere increases with depth, another confirmation that energy is flowing up from below.

8-1b The Chromosphere

Above the photosphere lies the chromosphere. Solar astronomers define the lower edge of the chromosphere as lying just above the visible surface of the Sun, with its upper regions blending gradually with the corona. The chromosphere is a layer with an irregular thickness on average less than Earth's diameter (Figure 8-1a). Because the chromosphere is roughly 1000 times fainter than the photosphere, you can see it with your unaided eyes only during a total solar eclipse when the Moon covers the brilliant photosphere. Then, the chromosphere flashes into view as a thin pink layer just above the photosphere. The word *chromosphere* comes from the Greek word *chroma,* meaning "color." The pink color is produced by the combined light of three bright emission lines—the red, blue, and violet Balmer lines of hydrogen (Section 7-3).

The chromosphere produces an emission spectrum, and Kirchhoff's second law tells you it therefore must be an excited, low-density gas. The chromosphere's density ranges from 10,000 times less dense than the air you breathe at the bottom of the chromosphere (near the photosphere) to 100 billion times less dense at the top (near the corona). Further analysis of solar spectra reveals that atoms in the lower chromosphere are ionized, and atoms in the higher layers of the chromosphere are even more highly ionized, having lost most or all of their electrons. Astronomers can find the temperature in different parts of the chromosphere from the amount of ionization. Just above the photosphere, the temperature falls to a minimum of about 4500 K and then rises rapidly (**Figure 8-3**) to the extremely high temperatures of the corona. The upper chromosphere is hot enough to emit X-rays and can be studied by X-ray telescopes in space (look back to Section 6-4c).

Solar astronomers can take advantage of the way spectral lines form to map the chromosphere. The gases of the chromosphere are transparent to nearly all wavelengths of visible light, but atoms in that gas are very good at absorbing photons of a few specific wavelengths. This produces some exceptionally strong dark absorption lines in the solar spectrum. A photon at one of those wavelengths is unlikely to escape from deeper layers to be received at Earth and can come to you only from higher in the Sun's atmosphere. A **filtergram** is an image of the Sun made using light only at the wavelength of one of those strong absorption lines such as the hydrogen Balmer series (Figure 2 in the Chapter 7 Concept Art: Atomic Spectra) to reveal detail in the upper regions of the chromosphere. Another way to study these layers of gas high in the Sun's atmosphere is to record solar images in the far-ultraviolet or in the X-ray part of the spectrum because those layers are very hot and emit most of their light at short wavelengths.

Figure 8-4a shows a filtergram made at the wavelength of the Balmer H-alpha line. **Figure 8-4b** shows complex structures called **spicules**, flamelike jets of gas extending upward into the chromosphere and lasting about 5 to 15 minutes. Seen at the limb of the Sun's disk, these spicules blend together and look like flames covering a burning prairie (**Figure 8-4c**), but they are more like the opposite of flames; spectra show that spicules are cooler gas from the lower chromosphere extending upward into hotter regions. Images at the center of the solar disk show that spicules spring up

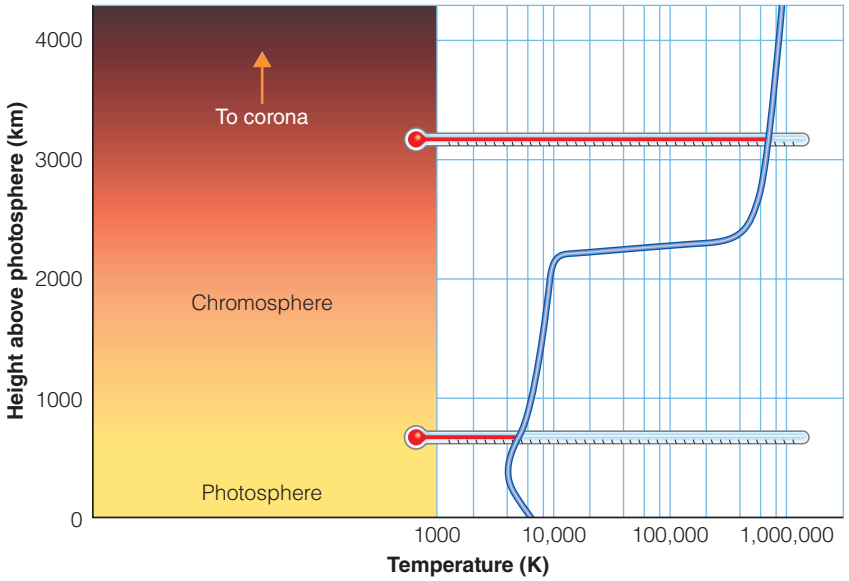
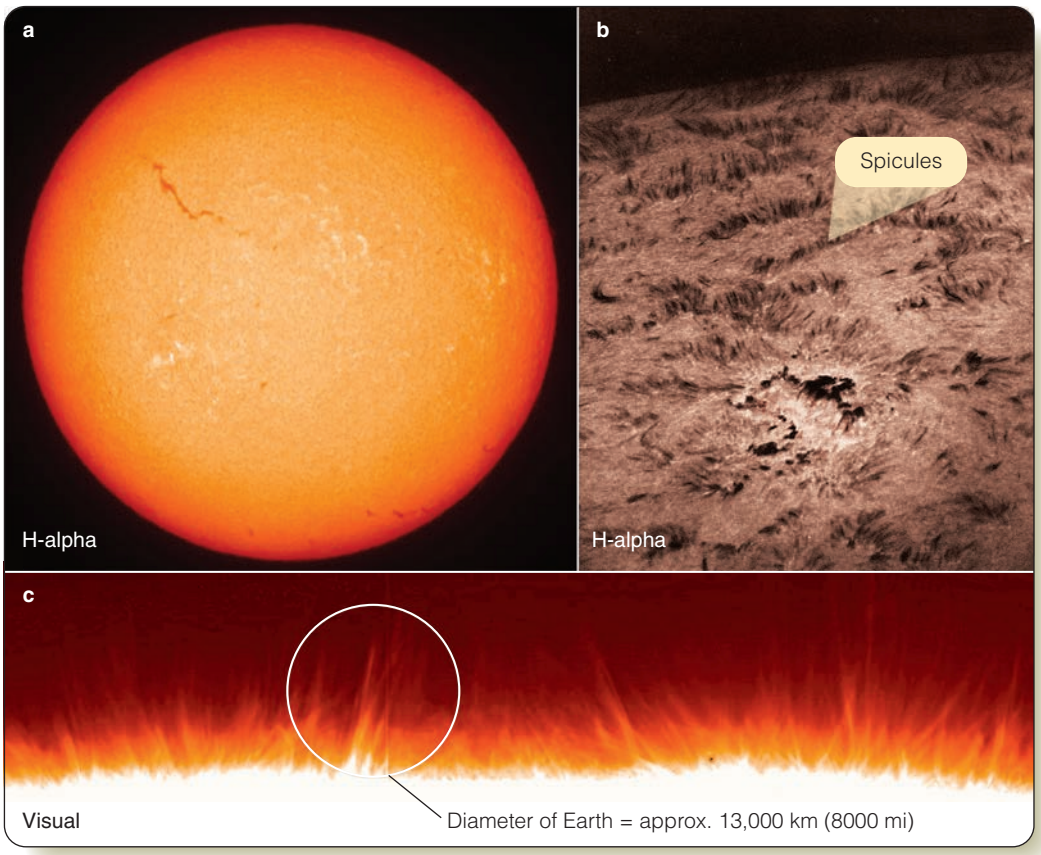


Figure 8-3 A plot showing the chromosphere's temperature profile. If you could place thermometers in the Sun's atmosphere, you would discover that the temperature increases from about 5800 K at the photosphere to 1 million K at the top of the chromosphere.

around the edge of supergranules like weeds around paving stones (Figure 8-4b).

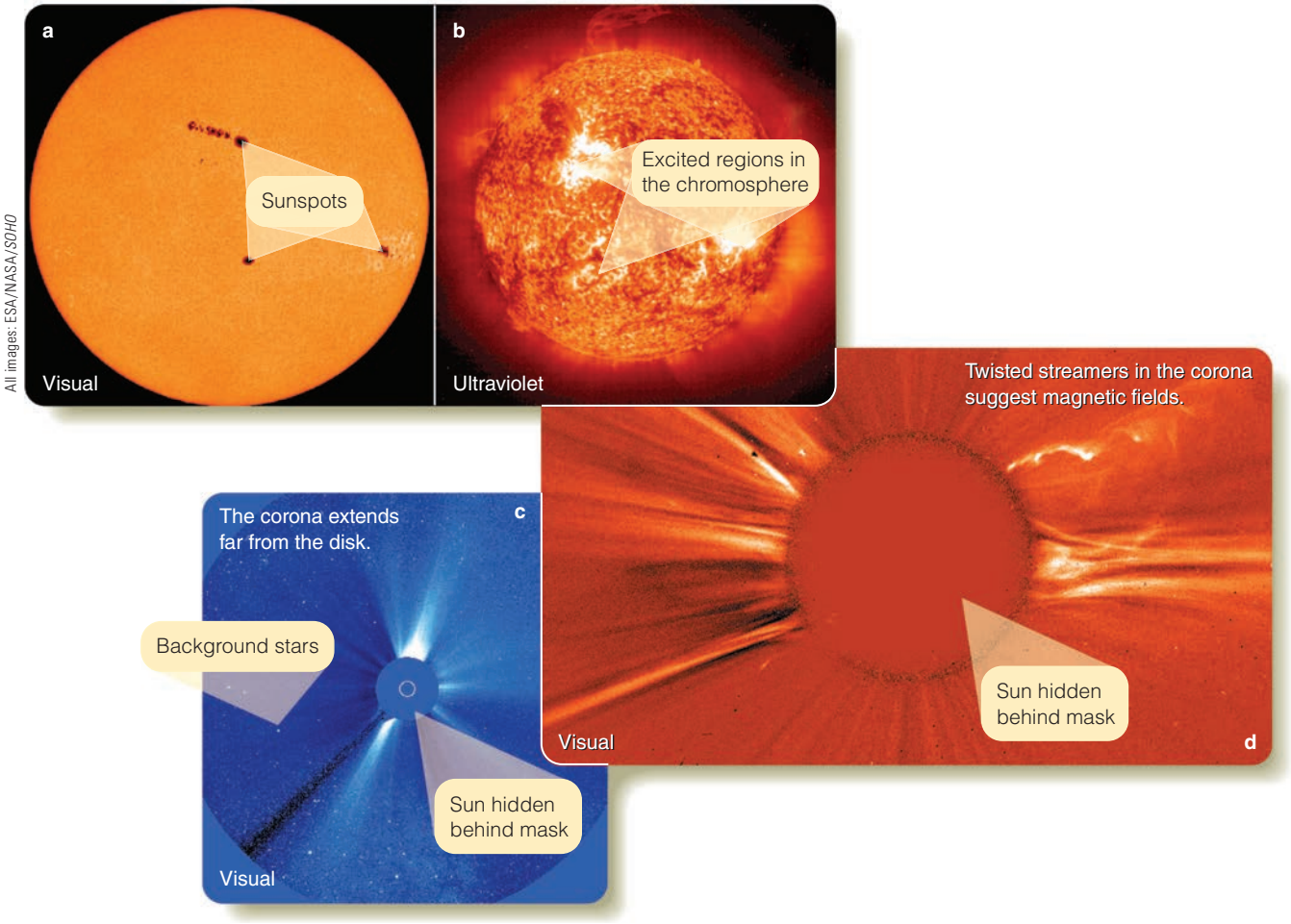
8-1c The Corona

The outermost part of the Sun's atmosphere is called the corona, after the Greek word for crown. The corona is so dim that, like the chromosphere, it is not visible in Earth's daytime sky because of the glare of scattered light from the Sun's brilliant photosphere. During a total solar eclipse, the innermost parts of the corona are visible to the unaided eye, as shown in Figure 8-1b (also, Figures 3-7 and 3-11). Observations made with specialized telescopes called **coronagraphs** can block the light of the photosphere and record the corona out beyond 20 solar radii, almost 10 percent of the way to Earth. Such images reveal streamers in



Panel a: BBSO; Panel b: © 1971 NSO/AURA/NSF; Panel c: JAXA/NASA/*Hinode*

Figure 8-4 (a) H-alpha filtergram of the Sun's disk. (b) H-alpha filtergrams reveal complex structures in the chromosphere that cannot be seen in ordinary visual-wavelength images, including spicules springing from the edges of supergranules. (c) Seen at the edge of the solar disk, spicules look like a burning prairie, but they are not at all related to burning. The white circle shows the size of Earth to scale. Compare with Figure 8-1a.

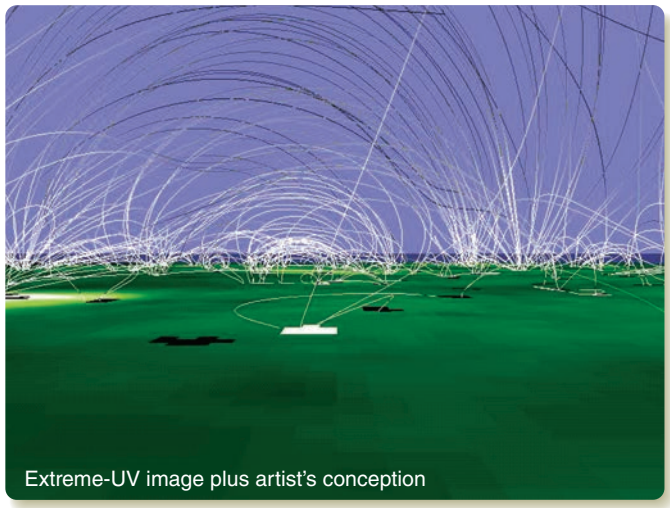


▲ **Figure 8-5** Images of the photosphere, chromosphere, and corona show relationships among the layers of the Sun's atmosphere. The visual-wavelength image in panel (a) was taken through a dense filter that produced the orange tint. (b) An ultraviolet image taken nearly simultaneously with image (a) shows hot regions in the chromosphere above the sunspots. (c) This image from the *Solar Max* spacecraft's coronagraph telescope emphasizes the solar corona by using an opaque mask to block the much brighter light from the Sun's photosphere. (d) Another coronagraph image shows complex structure in the inner corona.

the corona that follow magnetic lines of force in the Sun's magnetic field (**Figure 8-5**). Later in this chapter you will learn more about how features and activity in the Sun's atmosphere are controlled by magnetic fields.

The corona's spectrum, like that of the upper chromosphere, includes emission lines of highly ionized gases. In the lower corona, the atoms are not as highly ionized as they are at higher altitudes, and this tells you that the temperature of the corona rises with altitude. Just above the chromosphere the temperature is about 500,000 K, and in the outer corona the temperature can be 2 million K or more. The corona is hot enough to emit X-rays, but the coronal gas is not very luminous because its density is very low, with only 10^6 atoms/cm^3 in its lower regions. That is about 1000 trillion times less dense than the air you breathe. In its outer regions the corona contains only 1 to 10 atoms/cm^3, which is fewer than in the best vacuum in laboratories on Earth.

Astronomers continue to wonder how the corona and chromosphere can be so hot. Heat flows from hot regions to cool regions, never from cool to hot. So, how can the heat from the photosphere, with a temperature of only 5800 K, flow out into the much hotter chromosphere and corona? Observations made by the *Solar and Heliospheric Observatory (SOHO)* satellite have mapped a **magnetic carpet** of looped magnetic fields extending up through the photosphere (**Figure 8-6**); the largest loops could easily encircle Earth. Because the gas in the chromosphere and the corona is ionized and has very low densities, it can't resist being accelerated by movements of the magnetic fields. Turbulence below the photosphere seems to flick the magnetic loops back and forth and whip the gas about, heating it. Furthermore, observations by the *Hinode* spacecraft reveal magnetic waves generated by turbulence below the photosphere traveling up into the chromosphere and corona and heating the gas. In both cases, energy appears to flow out from the interior of the Sun to the chromosphere and corona not by radiation but by agitation of magnetic fields. The *Interface Region Imaging Spectrograph (IRIS)* space telescope was launched in 2013 to make rapid, high-resolution ultraviolet images of the upper



NASA/Stanford SOI/Lockheed-Martin ATC-SAL

▲ **Figure 8-6** Flying through the magnetic carpet. This computer model shows an extreme-ultraviolet image of a section of the Sun's lower corona (*green*) with black and white areas marking regions of opposite magnetic polarity. The model includes lines to show how the areas are linked by loops of magnetic force.

chromosphere and lower corona. Those data are helping advance investigations of the Sun's outer atmosphere and its puzzlingly high temperatures.

Ionized, low-density gas cannot cross magnetic fields, so in places where the Sun's field loops back toward the surface, the corona's gas is trapped in the vicinity of the Sun. However, some of the magnetic field lines are "open" and lead outward into space. At those locations the gas flows away from the Sun in the **solar wind** that can be considered an extension of the corona. The low-density gases of the solar wind blow past Earth at 300 to 800 km/s with gusts as high as 1000 km/s (more than 2 million mph).

8-1d The Solar Wind

Earth is bathed in the corona's hot breeze, which blows all the way to the outskirts of the Solar System. The *Voyager 1* and *2* spacecraft, launched in the 1970s to explore Jupiter and other outer planets, are now traveling through and investigating the region known as the **heliopause** where the solar wind collides with material in interstellar space. The *Interstellar Boundary Explorer (IBEX)* spacecraft, on the other hand, is able to analyze particles emitted from the heliopause without leaving Earth orbit.

Because of the solar wind, the Sun loses about 1 million tons per second, but that is only 10^{-14} of its total mass per year. Later in life, the Sun, like many other stars, will lose mass rapidly in a more powerful wind. You will see in future chapters how rapid outflowing winds affect the evolution of stars.

Do other stars have chromospheres, coronae (plural of corona), and stellar winds like the Sun? Stars are so far away they appear only as points of light even in the largest telescopes, but ultraviolet and X-ray observations suggest that the answer is yes, other stars have atmospheric features analogous to the Sun's. The spectra of many stars contain emission lines at far-ultraviolet wavelengths that could have formed only in the low-density, high-temperature gases of a chromosphere and corona. Also, many stars are sources of X-rays that seem to be produced by high-temperature gas in their chromospheres and coronae. This observational evidence gives astronomers good reason to consider the Sun to be a typical star, including all its complexity that can be seen from our nearby viewpoint.

8-1e Composition of the Sun

It seems as though it should be easy to learn the composition of the Sun just by studying its spectrum, but this is actually a difficult problem that wasn't well understood until the 1920s. The solution to that problem is part of the story of an important American astronomer who waited decades to get proper credit for her work.

In her 1925 PhD thesis, Cecilia Payne (**Figure 8-7**) invented the modern methods of interpreting spectra of the Sun and stars. For example, sodium lines are observed in the Sun's spectrum, so you can be sure that the Sun's atmosphere contains some sodium atoms. Payne came up with a mathematical procedure to determine just how many sodium atoms are there. She also proved that if spectral lines of a certain element are not detected in the Sun's spectrum, that element might still be present, but the gas is too hot or too cool, or the wrong density, for that type of atom to have electrons in the right energy levels to produce visible lines.

PF-(bygone1)/Alamy Stock Photo

▲ **Figure 8-7** Cecilia Payne (later, Payne-Gaposchkin; 1900–1979) was a British-born astrophysicist whose career in the United States included using mathematical techniques of quantum mechanics to determine that the true chemical composition of the Sun and other stars is mostly hydrogen and helium.

Payne's first calculations showed that more than 90 percent of the atoms in the Sun must be hydrogen and most of the rest are helium. In contrast, atoms like calcium, sodium, and iron that have strong lines in the Sun's spectrum are actually not very abundant. Rather, at the temperature of the Sun's atmosphere, those atoms are especially efficient at absorbing photons with wavelengths of visible light. At the time Payne did her original work, astronomers found it hard to believe her calculated abundances of hydrogen, helium, and other elements in the Sun. They especially found such a high abundance of helium unacceptable because helium lines are nearly invisible in the Sun's spectrum. Eminent astronomers dismissed Payne's results as obviously wrong; it was only several decades later that the scientific community realized the value of her work. Now we know that she was correct.

Abundances of elements in the Sun are presented in **Table 8-1**. Some of the abundances in the table, particularly of carbon, nitrogen, and oxygen (CNO), are subjects of ongoing controversy because of revised calculations of solar atmospheric conditions. Those calculations indicate that the CNO abundances might need to be lowered somewhat, but even so, the modern values are close to the ones determined in the 1920s by Payne.

Payne's work on the composition of the Sun illustrates the importance of fully understanding the interaction between light and matter in order to investigate objects in the Universe. In the next chapter, you will continue to follow the story of Cecilia Payne in regard to the main part of her thesis, which was interpreting the spectra of stars other than the Sun.

The layers of the solar atmosphere are all that astronomers can observe directly, but there are phenomena in those layers that reveal what it's like inside the Sun, your next destination.

8-1f Below the Photosphere

Almost no light emerges from below the photosphere, so you can't see into the solar interior. However, solar astronomers use a technique called **helioseismology**, analyzing naturally occurring vibrations in the Sun to explore its depths. Convective movements of gas in the Sun constantly produce vibrations—rumbles that would be much too low to hear with human ears even if your ears could survive a visit to the Sun's atmosphere. Some of these vibrations resonate in the Sun like sound waves in organ pipes. A vibration with a period of 5 minutes is strongest, but other vibrations have periods ranging from 3 to 20 minutes. These are very, very low-pitched sounds!

Astronomers can detect these vibrations by observing Doppler shifts in the solar surface. As a sound wave travels down into the Sun's interior, the changing density and temperature of the gas it moves through curves its path, and it returns to the surface. At the surface, it makes the photosphere heave up and down by small amounts—roughly plus or minus 15 km (10 mi). Multiple vibrations occurring simultaneously cover the surface of the Sun with a pattern of rising and falling regions that can be mapped using the Doppler effect (**Figure 8-8**). For example, the *SOHO* space telescope can observe solar oscillations continuously and is able to detect motions as slow as 1 mm/s (0.002 mph). Short-wavelength waves penetrate less deeply and travel shorter distances than longer-wavelength waves, so the vibrations of different wavelengths explore different layers in the Sun. Just as geologists can study Earth's interior by analyzing seismic waves from earthquakes, so solar astronomers can use helioseismology to explore the Sun's interior.

You can better understand how helioseismology works if you think of a duck pond. If you stood at the shore of a duck pond and looked down at the water, you would see ripples arriving from all parts of the pond. Because every duck on the pond contributes to the ripples, you could, in principle, study the ripples near the shore and draw a map showing the position and velocity of every duck on the pond. Of course, it would be difficult to untangle the different ripples. Nevertheless, all of the information would be there, lapping at the shore.

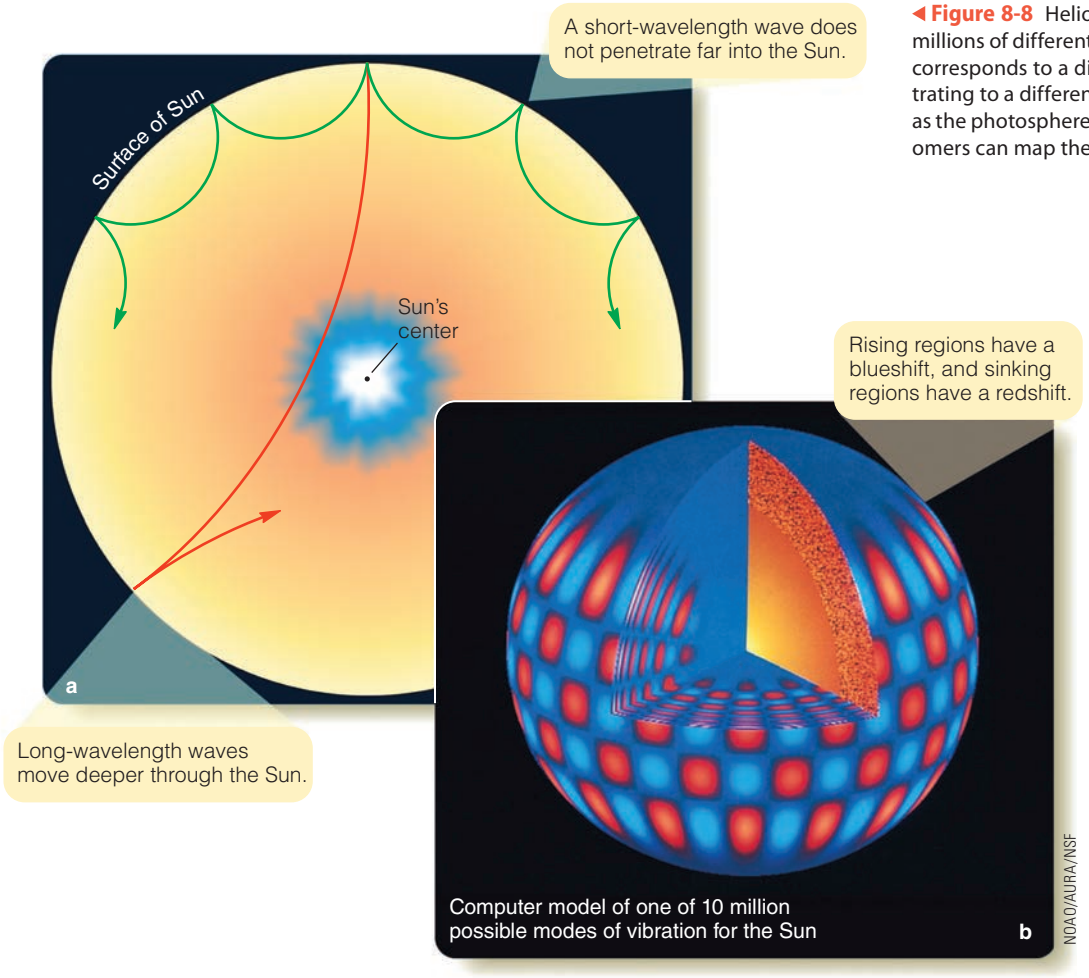
Helioseismology requires huge amounts of data, so astronomers have used a network of telescopes around the world operated by the Global Oscillation Network Group (GONG). The network observes the Sun continuously for weeks at a time as Earth rotates; in other words, the Sun never sets on GONG. Solar astronomers then use supercomputers to separate the different vibration patterns on the solar surface and determine the strength of the waves at many different wavelengths.

Helioseismology has allowed astronomers to map the temperature, density, and rate of rotation in the interior of the Sun, as well as find the positions and speeds of great currents of gas flowing below the photosphere. For example, the depth of the region of convection is now known to be exactly 29 percent of the radius of the Sun. That detailed information confirms a model developed to understand the cycles of solar activity that you will learn about in the next section.

Table 8-1 The Most Abundant Elements in the Solar Atmosphere*

Element	Atomic Number	Percentage by Number of Atoms	Percentage by Mass
Hydrogen	1	92.6	74.9
Helium	2	7.3	23.7
Carbon	6	0.023	0.22
Nitrogen	7	0.063	0.071
Oxygen	8	0.045	0.59
Neon	10	0.0069	0.11
Magnesium	12	0.0033	0.064
Silicon	14	0.0032	0.073
Sulfur	16	0.0014	0.037
Iron	26	0.0027	0.12

*Based on Lodders 2003, *Astrophysical Journal* 591, 1220.


A short-wavelength wave does not penetrate far into the Sun.
Surface of Sun
Sun's center
Rising regions have a blueshift, and sinking regions have a redshift.
a
Long-wavelength waves move deeper through the Sun.
Computer model of one of 10 million possible modes of vibration for the Sun
b
NOAO/AURA/NSF

◀ **Figure 8-8** Helioseismology: The Sun can vibrate in millions of different patterns or modes, and each mode corresponds to a different vibration wavelength penetrating to a different level. By measuring Doppler shifts as the photosphere moves gently up and down, astronomers can map the inside of the Sun.

Practicing **SCIENCE**

What evidence leads astronomers to conclude that temperature increases with height in the chromosphere and corona? In astronomy, as in any science, evidence is crucial, and gathering evidence means making observations and measurements.

Solar astronomers can observe the spectrum of the chromosphere, and they find that atoms there are more highly ionized (have lost more electrons) than atoms in the photosphere. Atoms in the corona are even more highly ionized. That must mean the chromosphere and corona are hotter than the photosphere.

A central part of practicing science is gathering, evaluating, and understanding evidence. Now, continue investigating the Sun by comparing it with other stars. ***What evidence leads astronomers to conclude that some stars have chromospheres and coronae like those of the Sun?***

8-2 Solar Activity

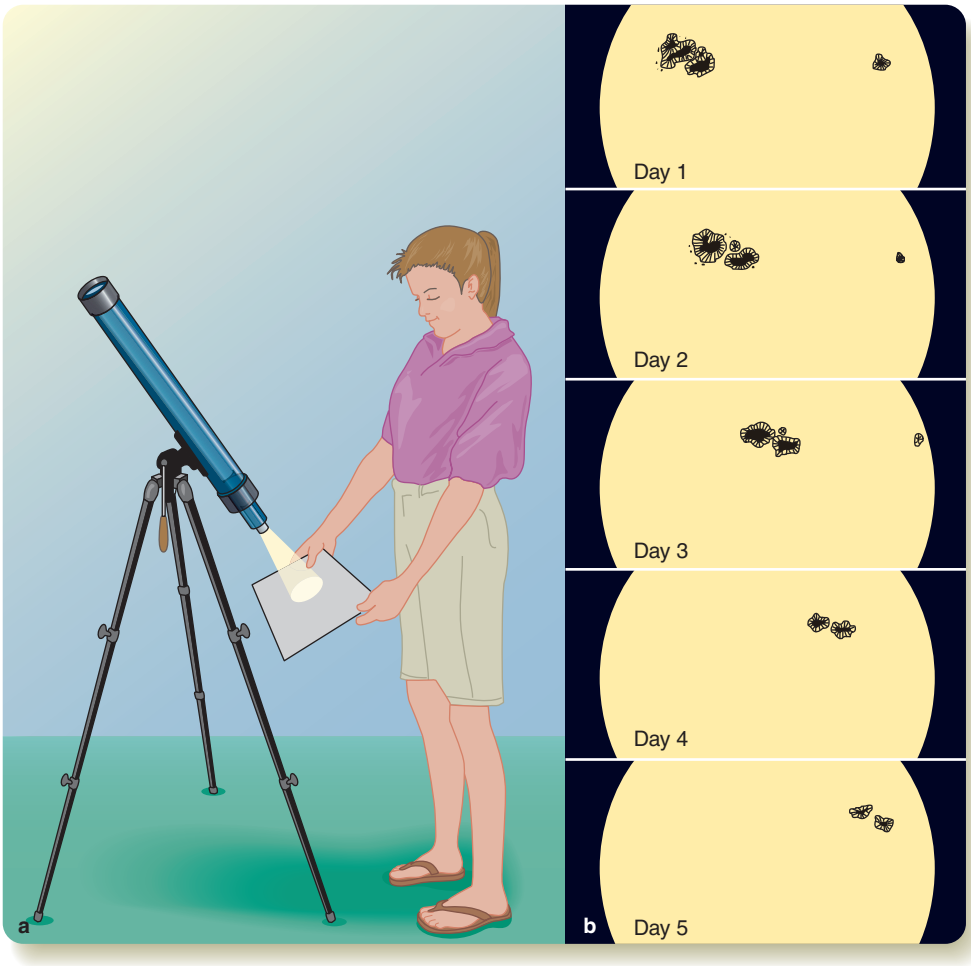
The Sun is not quiet. It has storms larger than Earth that last for weeks and unimaginably vast eruptions. All of these seemingly different forms of solar activity have one thing in common—magnetic fields. The weather on the Sun is magnetic.

8-2a Observing the Sun

Solar activity is often visible with even a small telescope, but you should be very careful if you try to observe the Sun. Sunlight is intense, and the infrared radiation in sunlight is especially dangerous because your eyes can't detect it. You don't sense how intense the infrared is, but it is converted to thermal energy in your eyes and can burn and scar your retinas.

It is not safe to look directly at the Sun, and it is even more dangerous to look at the Sun through any optical instrument such as a telescope, binoculars, or even the viewfinder of a camera. The light-gathering power of such an optical system concentrates the sunlight and can cause severe injury. Never look at the Sun with any optical instrument unless you are certain it is safe. **Figure 8-9a** shows a safe way to observe the Sun with a small telescope.

In the early 17th century, Galileo observed the Sun with a thick, dark filter over his telescope and saw spots on its surface. Day by day, he saw the spots moving across the Sun's disk and concluded that the Sun is a rotating sphere. If you repeated Galileo's observations, you would probably also detect **sunspots**, a view that would look something like **Figure 8-9b**.



◀ **Figure 8-9** (a) Looking through a telescope at the Sun is dangerous, but you can always view the Sun safely with a small telescope by projecting its image on a white screen. (b) If you sketch the location and structure of sunspots on successive days, you will see the rotation of the Sun and gradual changes in the size and structure of sunspots, just as Galileo did in 1612.

8-2b Sunspots

The dark sunspots that you see at visible wavelengths only hint at the complex processes that go on in the Sun's atmosphere. To explore those processes, you need to analyze images and spectra at a wide range of wavelengths.

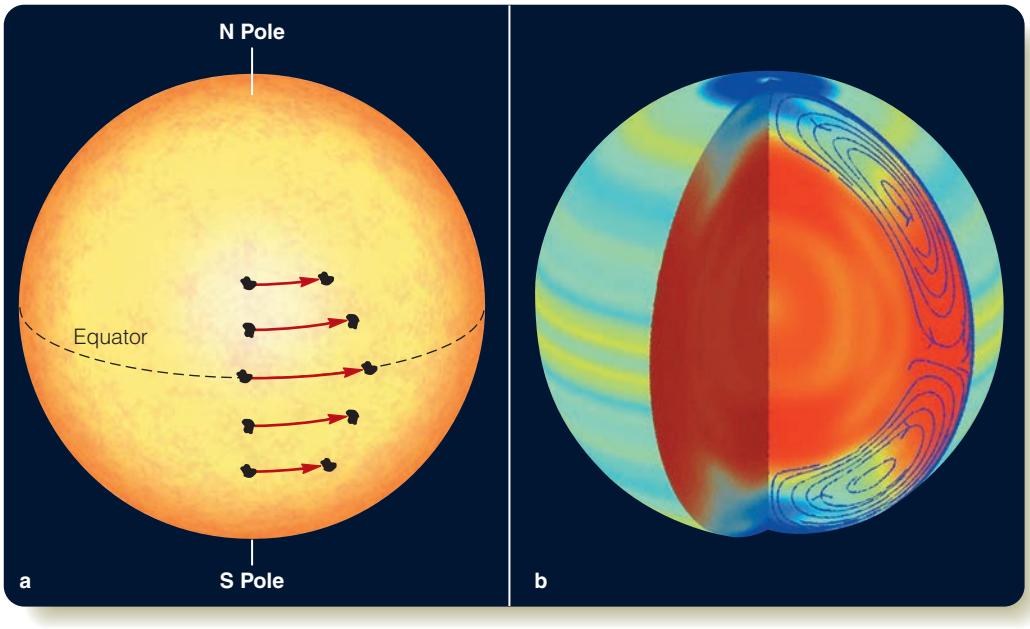
Study **Sunspots and the Solar Magnetic Cycle** and notice five important points and four new terms:

1. Sunspots are cool, relatively dark spots on the Sun's photosphere, usually appearing in groups, which form and disappear over time scales of weeks and months.
2. Sunspot numbers follow an 11-year cycle, becoming more numerous, reaching a maximum, and then becoming much less numerous. The *Maunder butterfly diagram* shows how the location of sunspots also changes during a cycle.
3. The *Zeeman effect* gives astronomers a way to measure the strength of magnetic fields on the Sun and provides evidence that sunspots contain, and are caused by, strong local magnetic fields. When the magnetic properties of sunspots are considered, the 11-year cycle is understood to be really a 22-year cycle.
4. The intensity of the sunspot cycle can vary from cycle to cycle and appears to have almost faded away during the *Maunder minimum* in the late 17th century. Some scientists hypothesize that this solar activity minimum was somehow connected with a significant cooling of Earth's climate that lasted for several centuries.
5. The evidence is clear that sunspots are parts of *active regions* dominated by magnetic fields that involve all layers of the Sun's atmosphere.

Sunspot groups are merely the visible traces of magnetically active regions. But what causes this magnetic activity? The answer is linked to a growth and decay cycle of the Sun's overall magnetic field.

8-2c The Sun's Magnetic Cycle

You are familiar with magnetic fields from classroom demonstrations with magnets and iron filings and from seeing the effect of Earth's magnetic field on a compass needle. The Sun's magnetic field is powered by the energy flowing outward through the moving currents of gas. The gas is highly ionized, so it is a very good conductor of electricity. When that electrically conducting gas rotates or is stirred by convection, some of the energy in the gas motion can be converted into magnetic field energy. This process is called the **dynamo effect**; it is understood to operate in Earth's core and produce Earth's



◀**Figure 8-10** (a) The photosphere of the Sun rotates faster at the equator than at higher latitudes. If you started five sunspots in a row along a north-south line, they would not stay lined up as the Sun rotates. (b) Detailed analysis of the Sun's rotation from helioseismology reveals that the interior of the Sun rotates differentially, with regions of relatively slow rotation (*blue*) and rapid rotation (*red*). Currents of gas are also detected moving from the equator toward the poles and back toward the equator.

magnetic field. Helioseismologists have found evidence that the dynamo effect generates the Sun's magnetic field at the bottom of the convection currents, deep below the photosphere.

Another important connection between solar gas motions and magnetic field lies in details of the Sun's rotation: The Sun does not rotate as a rigid body; this is possible because the Sun is entirely gas. For example, the equatorial region of the photosphere has a shorter rotation period than regions at higher latitudes (**Figure 8-10a**). At the equator, the photosphere rotates once every 24.5 days, but at latitude 45 degrees one rotation takes 27.8 days. This phenomenon is called **differential rotation**. Helioseismology maps of rotation in the Sun's interior (**Figure 8-10b**) reveal that the gas at different levels also rotates with different periods, another type of differential rotation. Both types of differential rotation, latitude-dependent and depth-dependent, seem to be involved in the Sun's magnetic cycle.

Although the magnetic cycle is not fully understood, the **Babcock model**, invented by astronomer Horace Babcock, explains the magnetic cycle as repeated tangling and untangling of the solar magnetic field. You have learned that an ionized gas is a very good conductor of electricity. This means that if the gas moves, embedded electrical currents and resulting magnetic fields must move with it. As a result, differential rotation drags the magnetic field along and wraps it around the Sun like a long string caught on a turning wheel. Rising and sinking convection currents then twist and concentrate the field into ropelike tubes. The Babcock model predicts that pairs of sunspots should occur where these tubes of concentrated magnetic energy burst through the Sun's surface (**Figure 8-11**).

Sunspots do tend to occur in groups or pairs, and the magnetic field around the pair resembles that around a bar magnet, with one end being magnetic north and the other end magnetic south. That is just what is expected if magnetic tubes, produced by convection and differential rotation according to the Babcock model, emerge from the Sun's surface through one sunspot in a pair and reenter through the other. At any one time, sunspot pairs south of the Sun's equator have reversed **polarity** (orientation of their magnetic poles) relative to those north of the Sun's equator. **Figure 8-12** illustrates this by showing sunspot pairs south of the Sun's equator moving with magnetic south poles leading, and sunspots north of the Sun's equator moving with magnetic north poles leading. At the end of an 11-year sunspot cycle, spots appear in the next cycle with reversed magnetic polarities relative to the spots in the previous cycle.

The Babcock model accounts for the reversal of the Sun's magnetic field from cycle to cycle. As the magnetic field becomes more and more tangled, adjacent regions of the Sun are dominated by magnetic fields that point in different directions. After years of tangling, the field becomes very complicated. Regions of weak north or south polarity "flip" into alignment with neighboring regions of stronger polarity. The entire field then quickly rearranges itself into a simpler pattern, the number of sunspots drops nearly to zero, and the cycle ends. Then, differential rotation and convection begin winding up the magnetic field to start a new cycle. The newly organized field is reversed relative to its predecessor, and the new sunspot cycle begins with the magnetic north end of sunspot groups replaced by magnetic south. Thus, although the solar activity cycle is 11 years long if you count numbers of sunspots, it is 22 years long if you consider sunspot magnetic field directions.

This magnetic cycle also explains the Maunder butterfly diagram. The Babcock model predicts that, as a sunspot cycle begins, the twisted tubes of magnetic force should first produce sunspot pairs at high latitudes on the Sun, exactly as is observed. In other words, the first sunspots in a new cycle appear far from the Sun's equator. Later in the cycle, when the field is more tightly wound, the tubes of magnetic force arch up through the

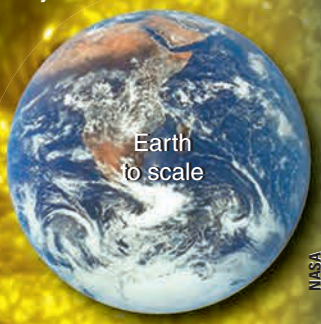
Sunspots and the Solar Magnetic Cycle

1 The dark spots that appear on the Sun are only the visible traces of complex regions of activity. Evidence gathered over many years and at a wide range of wavelengths shows that sunspots are clearly linked to the Sun's magnetic field.

Spectra show that sunspots are cooler than the photosphere with a temperature of about 4200 K. The photosphere has an average temperature of about 5800 K. Because the total amount of energy radiated by a surface depends on its temperature raised to the fourth power, sunspots look dark in comparison with the photosphere. Actually, a sunspot emits quite a bit of radiation. If the Sun were removed and only an average-size sunspot were left behind, it would be brighter than a full moon.

Visual

A typical sunspot is about twice the size of Earth, but there is a wide range of sizes. Sunspots appear, last for a few weeks to a few months, and then shrink away. Usually, sunspots occur in pairs or complex groups.



NASA

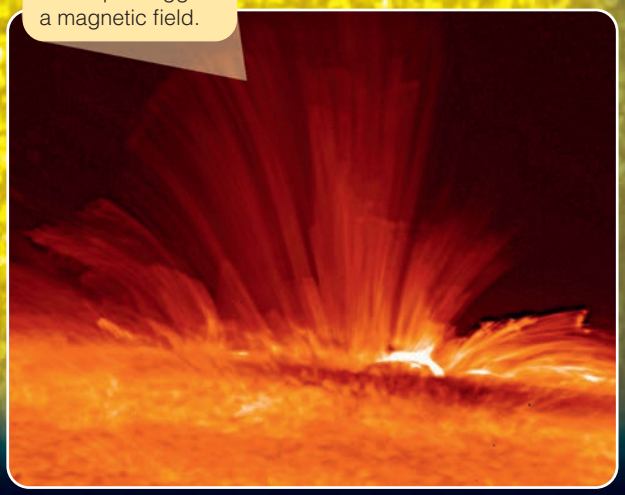
Umbra

Penumbra

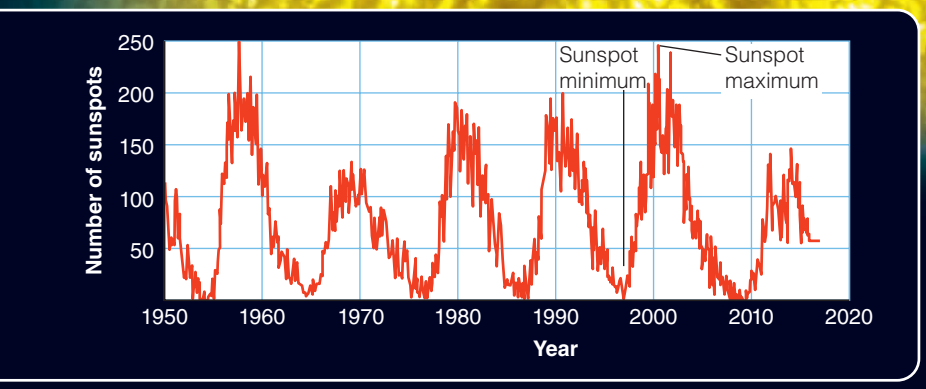
Sunspots are not shadows, but astronomers refer to the dark core of a sunspot as its *umbra* and the outer, lighter region as the *penumbra*.

JAXA/NASA/*Hinode*

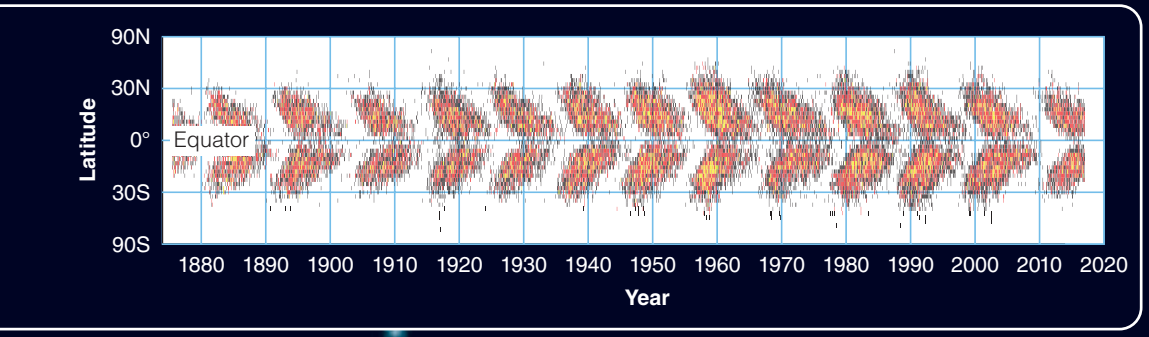
Streamers above a sunspot suggest a magnetic field.



JAXA/NASA/*Hinode*



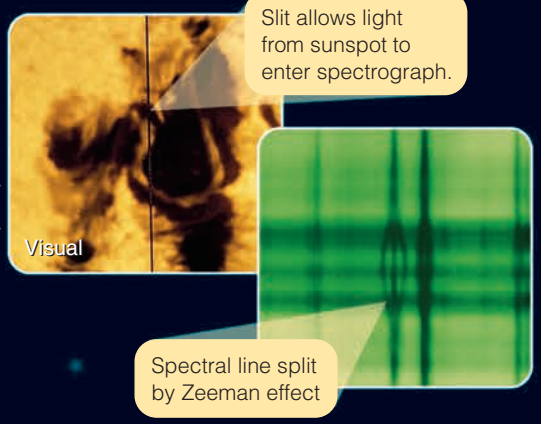
2 The number of spots visible on the Sun varies in a cycle with a period of 11 years. At maximum, there are often more than 100 spots visible. At minimum, there are very few or zero.



NASA MSFC/D. Hathaway

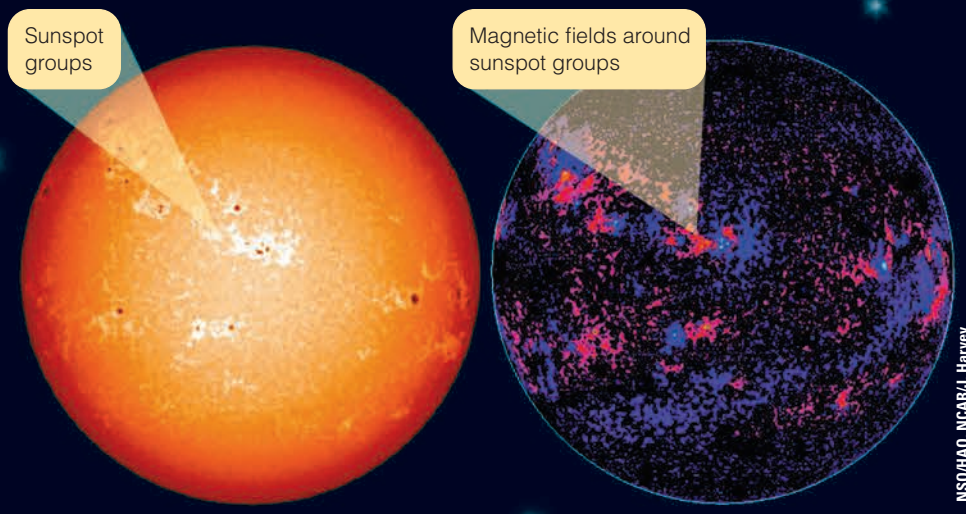
2a Early in a cycle, spots appear at high latitudes north and south of the Sun's equator. Later in the cycle, new spots appear closer to the Sun's equator. If you plot the latitude of sunspots versus time, the graph looks like butterfly wings, as shown in this **Maunder butterfly diagram**, named after E. Walter Maunder of Greenwich Observatory.

3 Astronomers can measure magnetic fields on the Sun using the **Zeeman effect**, as shown below. When an atom is in a magnetic field, the electron energy levels are altered, and the atom is able to absorb photons with a greater variety of wavelengths than the same atom not in a magnetic field. In this spectrum you see single spectral lines split into multiple components, with the separation between the components proportional to the strength of the magnetic field.



NSO/AURA/NSF

NSO/AURA/NSF



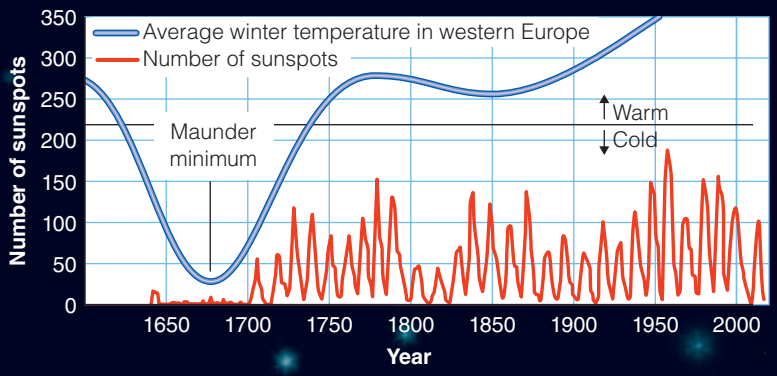
NSO/HAO, NCAR/J. Harvey

Ultraviolet filtergram

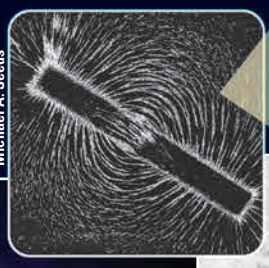
Magnetic image

Simultaneous images

3a Images of the Sun above show that sunspots contain magnetic fields a few thousand times stronger than Earth's. Such strong fields inhibit motions of ionized gas below the photosphere; consequently, convection is reduced below the sunspot, less energy is transported from the interior, and the Sun's surface at the position of the spot is cooler. Heat that is prevented from emerging at the sunspot's position is deflected and emerges around the sunspot, making the surrounding area hotter than the average photosphere. The deflected heat can be detected in ultraviolet and infrared images; the result is that the entire active region, including the sunspots, is actually emitting more energy than the same area of normal photosphere.

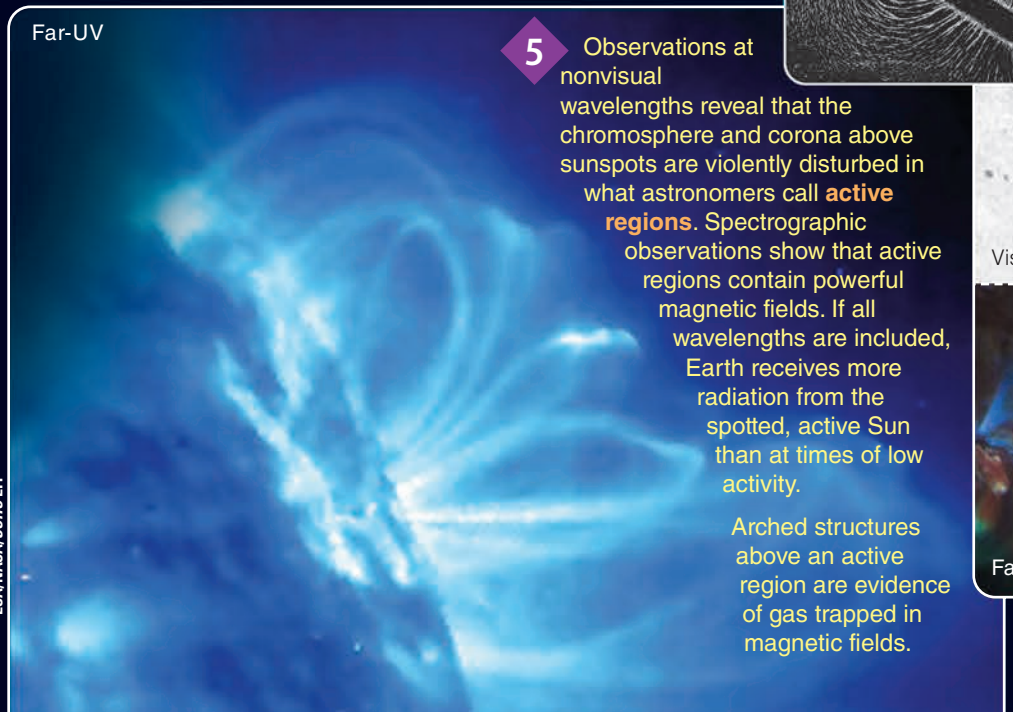


4 Historical records show that there were very few sunspots from about 1645 to 1715, a phenomenon known as the **Maunder minimum**. This coincides with the middle of a period called the "Little Ice Age," a time of unusually cool weather in Europe and North America from about 1500 to about 1850, as indicated by the blue line in the graph at left. Other such periods of cooler climate are known. Evidence suggests that there is a link between solar activity and the amount of solar energy Earth receives. This link has been confirmed by measurements made by spacecraft above Earth's atmosphere.



Michael A. Seeds

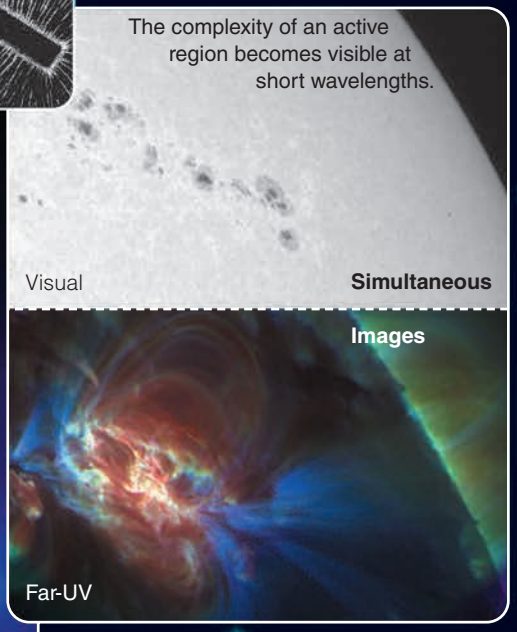
Magnetic fields can reveal themselves by their shape. For example, iron filings sprinkled over a bar magnet reveal an arched shape.



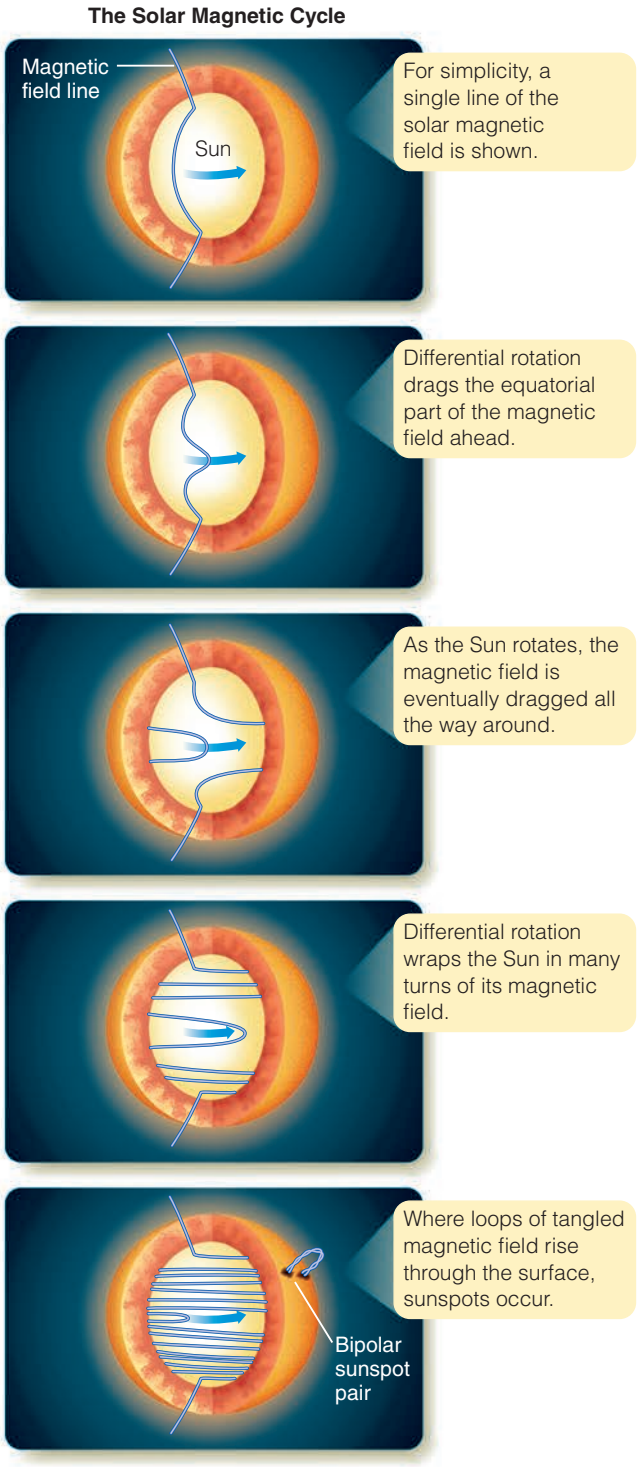
ESA/NASA/SOHO EIT

5 Observations at nonvisual wavelengths reveal that the chromosphere and corona above sunspots are violently disturbed in what astronomers call **active regions**. Spectrographic observations show that active regions contain powerful magnetic fields. If all wavelengths are included, Earth receives more radiation from the spotted, active Sun than at times of low activity.

Arched structures above an active region are evidence of gas trapped in magnetic fields.



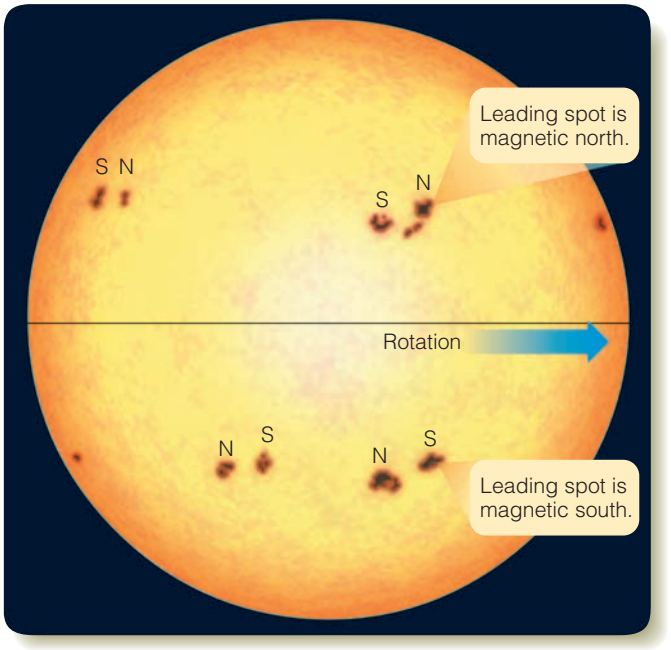
NASA/TRACE


The Solar Magnetic Cycle
Magnetic field line
Sun
For simplicity, a single line of the solar magnetic field is shown.
Differential rotation drags the equatorial part of the magnetic field ahead.
As the Sun rotates, the magnetic field is eventually dragged all the way around.
Differential rotation wraps the Sun in many turns of its magnetic field.
Where loops of tangled magnetic field rise through the surface, sunspots occur.
Bipolar sunspot pair

▲ **Figure 8-11** The Babcock model of the solar magnetic cycle explains the sunspot cycle as primarily a consequence of the Sun's differential rotation gradually winding up and tangling the magnetic field near the base of the Sun's outer, convective layer.

surface at lower latitudes. As a result, sunspot pairs later in a cycle appear closer to the equator.

A refinement of the Babcock model includes the **meridional flow**, which involves slow movements of gas from the Sun's equator to each pole and back, tens of thousands of kilometers below the



▲ **Figure 8-12** In sunspot groups, here simplified into pairs of major spots, the leading spot and the trailing spot have opposite magnetic polarity. Spot pairs in the Southern Hemisphere have reversed polarity from those in the Northern Hemisphere.

photosphere, that are detected by helioseismology measurements (Figure 8-10b). The meridional flow carries magnetic field bundles toward the poles from active regions at lower latitudes during each sunspot cycle, thereby establishing the foundation of the next cycle's magnetic field.

Notice the power of a scientific model. Even though the model of the sky in Section 2-2a and the model of atoms in Section 7-1a are only partially correct, they serve as organizing themes to guide further exploration. Similarly, many of the details of the solar magnetic cycle are not yet understood. The Babcock model may be partly incorrect or incomplete. For example, the start of the solar cycle that should have begun around mid-2008 was delayed by 15 months to late 2009. The subsequent cycle maximum in 2013 was the weakest in 100 years in terms of the number of sunspots and amount of solar magnetic activity. The simple Babcock model does not easily account for such an anomaly. Nevertheless, the model provides a framework around which to organize descriptions and investigations of complex solar activity (**How Do We Know? 8-1**).

8-2d Chromospheric and Coronal Activity

The solar magnetic fields extend high into the chromosphere and corona, where they produce beautiful and powerful phenomena. Study **Solar Activity and the Sun–Earth Connection** and notice three important points and seven new terms:

1. All solar activity is magnetic. The arched shapes of *prominences* are produced by magnetic fields, and *filaments* are prominences seen from above.

HOW DO WE KNOW? 8-1

Confirmation and Consolidation

What do scientists do all day?

The scientific method is sometimes portrayed as a kind of assembly line where scientists crank out new hypotheses and then test them through observation. In reality, scientists don't often generate entirely new hypotheses. And it is rare that an astronomer makes an observation that disproves a long-held theory and triggers a revolution in science. Then what is the daily grind of science really about?

Many observations and experiments confirm already-tested hypotheses. The biologist knows that all worker bees in a hive are sisters because they are all female, and they all had the same mother, the queen bee. A biologist can study the DNA from many workers and confirm that hypothesis. By repeatedly checking and thereby *confirming* a hypothesis, scientists build confidence in the hypothesis and may be able to extend it. Do all of the workers in a hive have the same father, or did the queen mate with more than one male drone?

Another aspect of routine science is *consolidation,* the linking of a hypothesis to other well-studied phenomena. A biologist can study yellow jacket wasps from a single nest and discover that the wasps, too, are sisters. There must be a queen wasp who lays all of the eggs in a nest. But, in a few nests, the scientist may find two sets of unrelated sister workers. Those nests evidently contain two queens sharing the nest for convenience and protection. From his study of wasps, the biologist consolidates what he knows about bees with what others have learned about wasps and reveals something new: that bees and wasps have evolved in similar ways for similar reasons.

Confirmation and consolidation allow scientists to build confidence in their understanding and extend it to explain more about nature.

Michael Durham/Minden Pictures/Getty Images

A yellow jacket is a wasp from a nest containing a queen wasp.

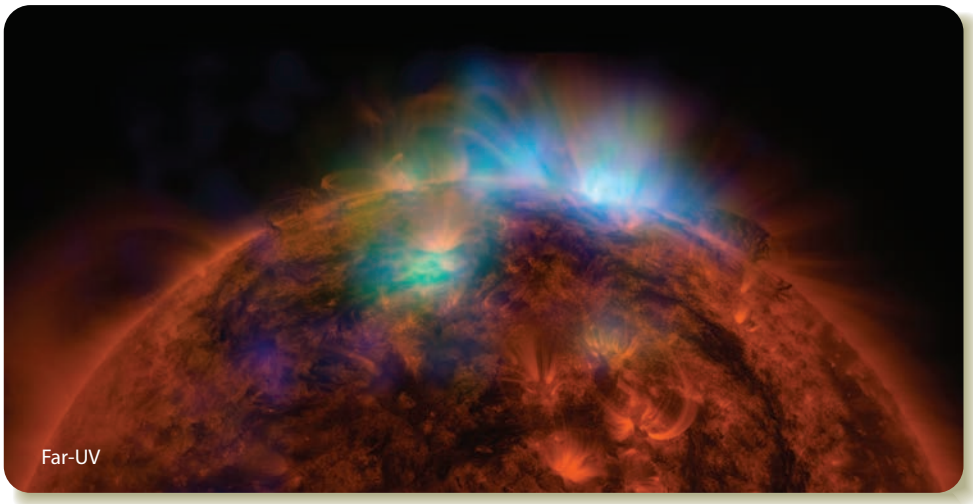
2 Tremendous amounts of energy can be stored in arches of magnetic fields, and when two arches encounter each other, a *reconnection event* can cause powerful eruptions called *flares.* Although these eruptions occur far from Earth, they can affect us in dramatic ways. For example, *coronal mass ejections (CMEs)* can trigger communications blackouts and *auroras.*

3 In some regions of the solar surface, the magnetic field does not loop back. High-energy gas from these *coronal holes* flows outward and produces much of the solar wind.

Images of the active Sun often show eruptive prominences, enormous arches that are shaped by magnetic fields and have sizes that dwarf Earth, standing above active regions on the solar limb (**Figure 8-13**).

Auroras are sometimes called the "northern lights," but they can be viewed often from high latitudes in both the Northern and Southern Hemisphere. Now, if you ever have an opportunity to watch a beautiful aurora display, you will know that you are actually seeing spectral emission lines from gases in Earth's upper atmosphere excited to glow by a complicated interaction with the solar wind and Earth's magnetic field (look again at the right-hand page of the Concept Art: Solar Activity and the Sun–Earth Connection).

A series of solar eruptions in August–September 1859 produced electromagnetic disturbances at Earth's surface so severe that telegraph equipment caught on fire and operators



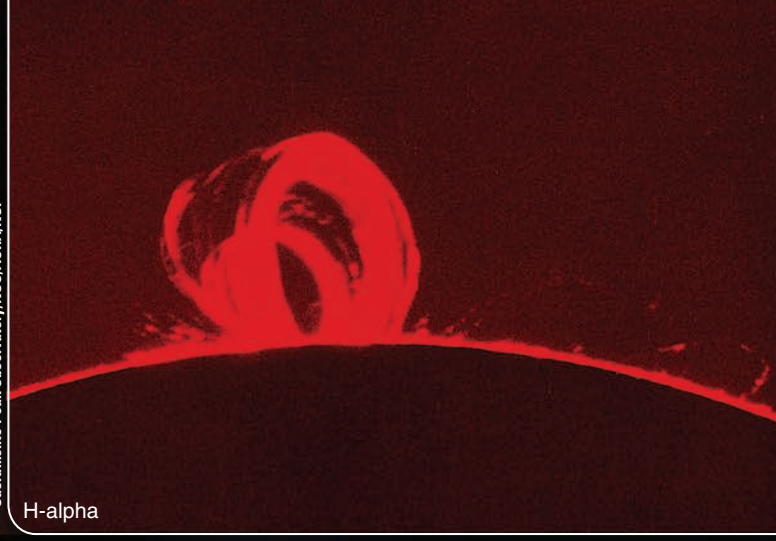
NASA/JPL-Caltech/GSFC

◀**Figure 8-13** X-ray image (*green* and *blue*) of the solar atmosphere above an active region taken by the *NuSTAR* space telescope, combined with an extreme UV image (*red* and *orange*) from the *Solar Dynamics Observatory (SDO)* spacecraft. The X-ray emission is from gas at temperatures of more than 3 million K, whereas the UV emission is from about 1 million K.

Solar Activity and the Sun–Earth Connection

1 Magnetic phenomena in the chromosphere and corona, like magnetic weather, result as constantly changing magnetic fields in the Sun's atmosphere trap ionized gas to produce beautiful arches and powerful outbursts. Some of this solar activity can affect Earth's magnetic field and atmosphere.

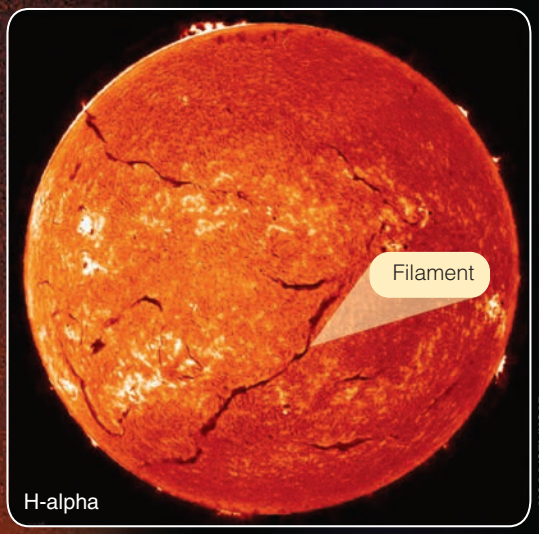
This ultraviolet image of the solar surface was made by the NASA *TRACE* spacecraft. It shows hot gas trapped in magnetic arches extending above active regions. At visual wavelengths, you would see sunspot groups in these active regions.



Sacramento Peak Observatory/NSO/AURA/NSF

H-alpha

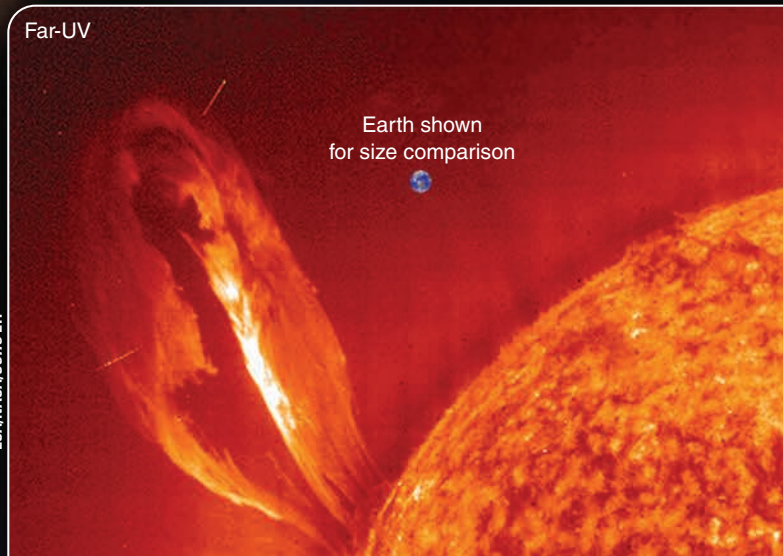
1a A **prominence** is composed of ionized gas trapped in a magnetic arch rising up through the photosphere and chromosphere into the lower corona. Seen during total solar eclipses at the edge of the solar disk, prominences look pink because of emission in the three hydrogen Balmer lines, H-alpha, H-beta, and H-gamma. The image above shows the arch shape indicating the influence of magnetic fields. Seen from above against the Sun's bright surface, prominences form dark **filaments**, shown in the photo below.



H-alpha

NOAA/SEL/USAF

1b Quiescent prominences may hang in the lower corona for many days, whereas eruptive prominences burst upward in hours. The eruptive prominence below is many Earth diameters in length.



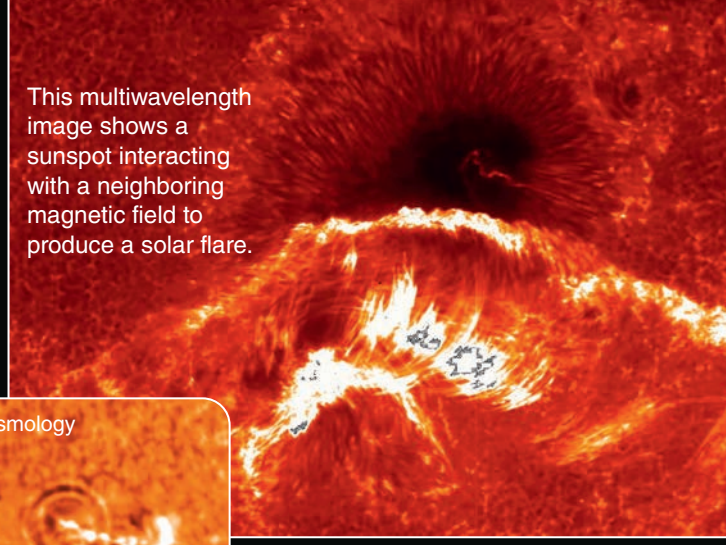
ESA/NASA/*SOHO* EIT

The gas in prominences may be 60,000 to 80,000 K, quite cold compared with the low-density gas in the corona, which may be as hot as a million K.

TRACE/NASA

Ultraviolet

2 Solar **flares** rise to maximum in minutes and decay in an hour. They occur in active regions where oppositely directed magnetic fields meet and cancel each other in what are called **reconnection events**. Energy stored in the magnetic fields is released as short-wavelength photons plus high-energy protons and electrons. X-ray and ultraviolet photons reach Earth in 8 minutes and increase ionization in our upper atmosphere, which can interfere with radio communications. Particles from flares reach Earth hours or days later as gusts in the solar wind, which can distort Earth's magnetic field and disrupt navigation systems. Solar flares can also cause surges in electrical power lines and damage to Earth satellites.



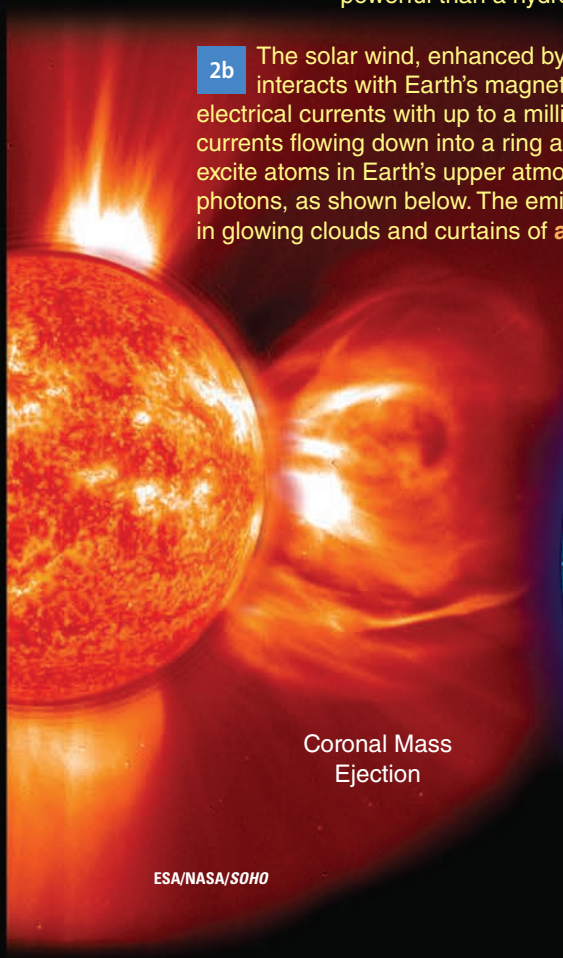
This multiwavelength image shows a sunspot interacting with a neighboring magnetic field to produce a solar flare.

JAXA/NASA/*Hinode*

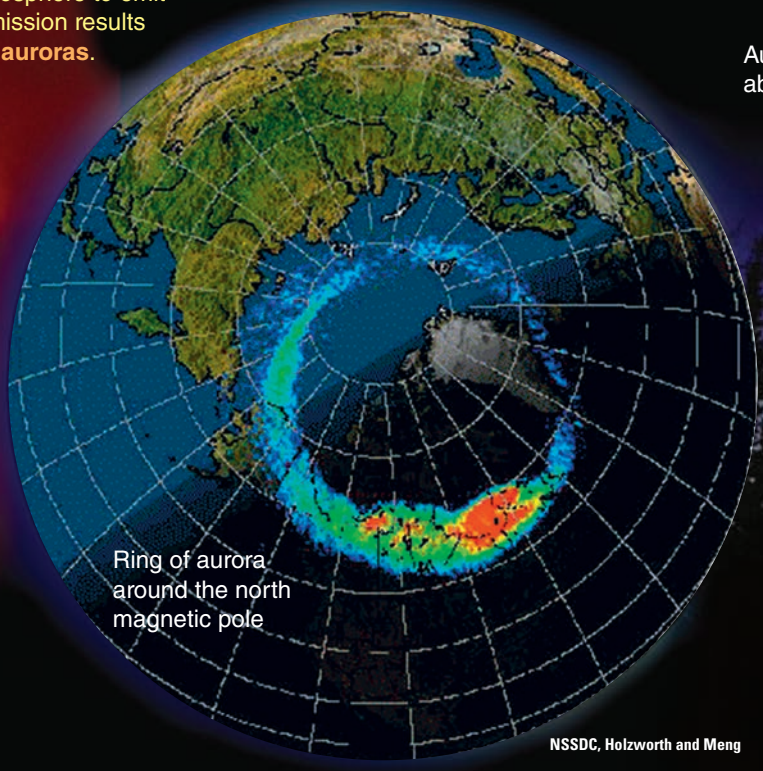
2a At right, waves rush outward at 50 km/s from the site of a solar flare 40,000 times stronger than the 1906 San Francisco earthquake. The biggest solar flares can be a billion times more powerful than a hydrogen bomb.

ESA/NASA/*SOHO* MDI

2b The solar wind, enhanced by eruptions on the Sun, interacts with Earth's magnetic field and can create electrical currents with up to a million megawatts of power. Those currents flowing down into a ring around Earth's magnetic poles excite atoms in Earth's upper atmosphere to emit photons, as shown below. The emission results in glowing clouds and curtains of **auroras**.



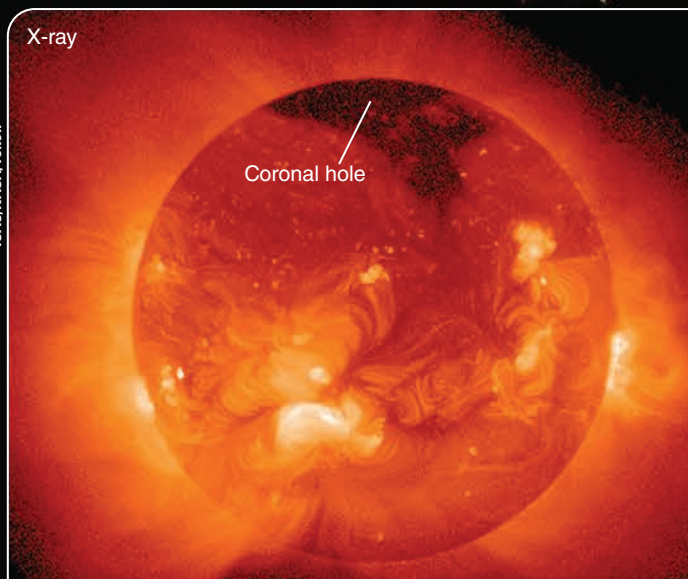
ESA/NASA/*SOHO*



NSSDC, Holzworth and Meng

Auroras occur about 130 km above Earth's surface.

2c Reconnection events can release enough energy to blow large amounts of ionized gas outward from the corona in **coronal mass ejections (CMEs)**. If a CME strikes Earth, it can produce especially violent disturbances in Earth's magnetic field.



ISAS/NASA/*Yohkoh*

3 Much of the solar wind comes from **coronal holes** where the magnetic field does not loop back into the Sun. These open magnetic fields allow ionized gas in the corona to flow away as the solar wind. The dark area in the X-ray image at right is a coronal hole.

received painful electrical shocks. A 2010 study by the Metatech Corporation (funded by NASA) indicated that if a solar eruption as large as the ones in 1859 occurred today, it would produce electrical blackouts affecting 40 percent of U.S. households that could last for months, waiting for destroyed electrical transmission and generation equipment to be replaced. There was a solar eruption of that magnitude in 2010, but the active region was pointed mostly away from Earth. Humanity became fully aware of the size of that event only because of data from a combination of the *Solar Dynamics Observatory (SDO)* and the twin *Solar TErrestrial RElations Observatory (STEREO)* spacecraft monitoring the Sun from different positions in the Solar System. You can imagine that scientists would like to identify early warning signs that such an eruption is imminent, as well as strategies to protect the electrical and communications equipment on which human civilization is now so dependent.

8-2e The Solar Constant

Even a small change in the Sun's energy output could produce dramatic changes in Earth's climate, but humanity knows very little about long-term variations in the Sun's energy output. The energy production of the Sun can be monitored by measuring the amount of solar radiation reaching Earth. Of course, you should include all wavelengths of electromagnetic radiation from X-rays to radio waves, so you need to correct for absorption by Earth's atmosphere or make the measurement from space. The result, called the **solar constant**, amounts to about 1370 J/s/m^2 (equivalent to W/m^2). But is the Sun really constant?

Measurements by the *Solar Maximum Mission* satellite showed variations in the energy received from the Sun by about 0.1 percent that last for days, weeks, or years, including one pattern of variation that appears to be timed with the magnetic activity cycle. Superimposed on both the random and cyclical variations is a very slight long-term decrease of about 0.018 percent per year that has been confirmed by observations made with other instruments. This long-term decrease may be related to a cycle of activity on the Sun with a period longer than the 22-year magnetic cycle. Thus, careful measurements show that the solar constant is not really constant.

As you learned previously, the "Little Ice Age" was a period of unusually cool weather in Europe and America that lasted from about 1500 to 1850. The average temperature worldwide was about 1°C cooler than it is now. This period of cool weather corresponded very roughly to the Maunder minimum, a period of reduced solar activity—few sunspots and auroral displays and little or no corona visible during solar eclipses. Scientists do not yet completely understand how those changes in the Sun's surface activity would connect to changes in Earth's average temperature. The measured changes in the modern solar "constant" seemingly would cause Earth to become slightly cooler, yet our planet is clearly observed to be warming. In a later chapter, you will learn more about the complex interaction between solar input, human activity, and changes in Earth's climate.

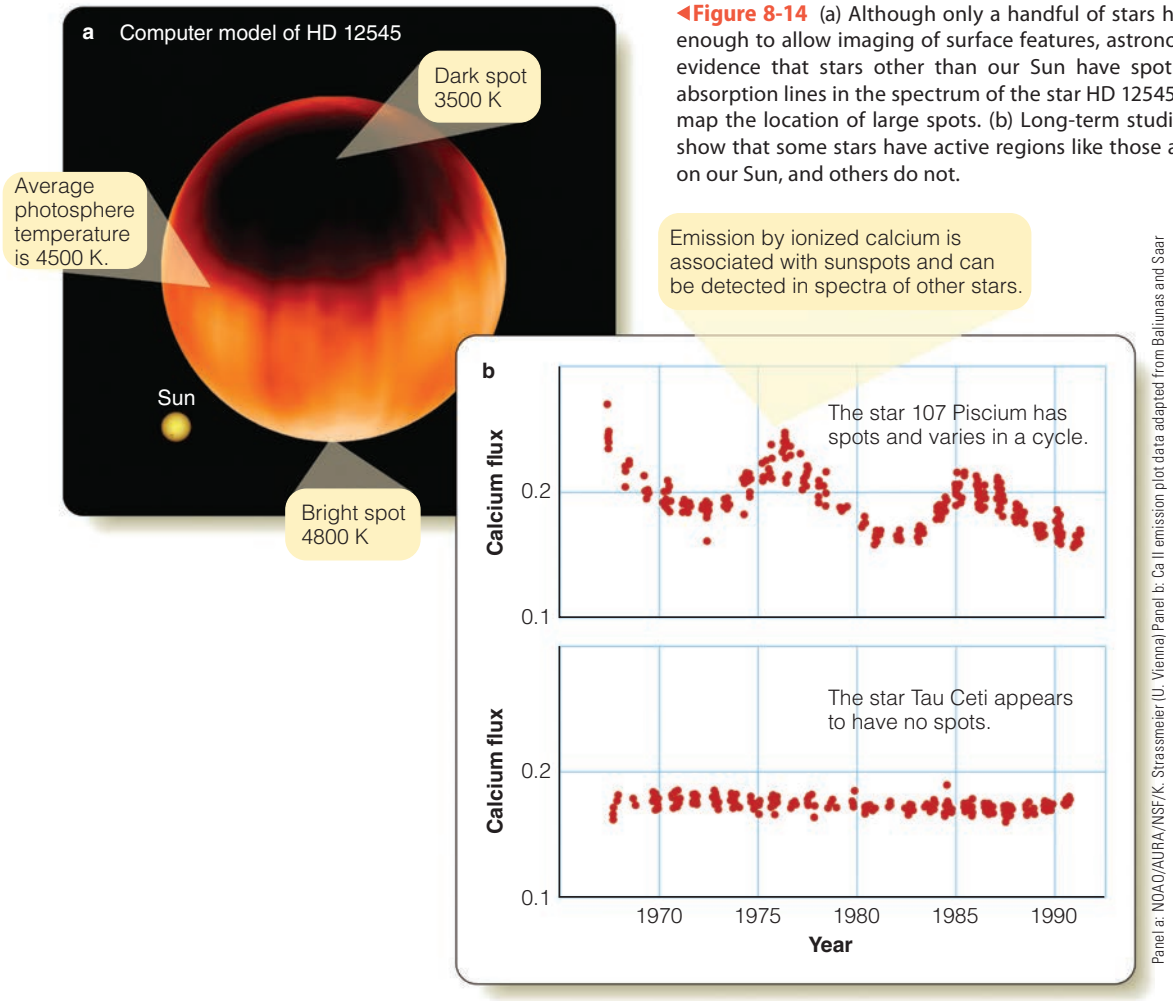
8-2f Spots and Magnetic Cycles of Other Stars

The Sun seems to be a representative star, so you should expect other stars to have cycles of "starspots" and magnetic activity similar to the Sun's. This is difficult to demonstrate observationally because, with few exceptions, the stars are too small or too far away to allow detection of surface detail. Some stars, however, vary in brightness in ways that suggest that they are covered by dark spots. As these stars rotate, their total brightness changes slightly, depending on the number of spots on the side facing Earth. High-precision spectroscopic analysis has even allowed astronomers to map the locations of spots on the surfaces of certain stars (**Figure 8-14a**). Such results confirm that the sunspots seen on our Sun are not unusual.

Certain features found in stellar spectra might be associated with magnetic fields by analogy with the Sun. Regions of strong magnetic fields on the solar surface emit strongly at the central wavelengths of the two strongest lines of ionized calcium. This calcium emission appears in the spectra of other Sun-like stars and suggests that these stars, too, have strong magnetic fields at some locations on their surfaces. In some cases, the strength of this calcium emission varies over periods of days or weeks and indicates that the stars have active regions and are rotating with periods similar to those of the Sun. These stars presumably have "starspots" as well.

In 1966, astronomers at Mt. Wilson observatory began a long-term project that monitored the strengths of these calcium emission features in the spectra of 91 stars with photosphere temperatures ranging from 1000 K hotter than the Sun to 3000 K cooler that were considered most likely to have Sun-like magnetic activity on their surfaces. The observations show that the strength of the calcium emission varies over periods of years. The calcium emission averaged over the Sun's disk varies with the sunspot cycle, and similar periodic variations can be seen in the spectra of some of the stars studied (**Figure 8-14b**). The star 107 Piscium, for example, appears to have a starspot cycle lasting nine years. At least one star, Tau Boötis, has been observed to reverse its magnetic field. This kind of evidence shows that stars like the Sun have similar magnetic cycles, and that the Sun is normal in this respect.

It is interesting to note that 15 percent of the Sun-like stars in the Mt. Wilson study were found to have very low activity levels; some astronomers have speculated that those stars are in phases equivalent to the Sun's Maunder minimum. On the other hand, investigators using the *Kepler* space telescope discovered that 148 solar-type stars from a sample of 80,000 had "superflares" more than 100 times more powerful than the largest solar flare ever observed.



Panel a: NOAO/AURA/NSF/K. Strassmeier (U. Vienna) Panel b: Ca II emission plot data adapted from Baliunas and Saar

◀**Figure 8-14** (a) Although only a handful of stars have angular sizes large enough to allow imaging of surface features, astronomers have found clear evidence that stars other than our Sun have spots. Detailed analysis of absorption lines in the spectrum of the star HD 12545 allows astronomers to map the location of large spots. (b) Long-term studies of calcium emission show that some stars have active regions like those around sunspot groups on our Sun, and others do not.

Practicing **SCIENCE**

What kind of activity would the Sun have if it didn't rotate differentially? Imagining a physical system with one factor changed is something scientists do to help them understand a concept.

Consider the Babcock model for solar magnetic activity cycles. If the Sun didn't rotate differentially, with its equator turning in a shorter period than its higher latitudes, then the magnetic field would not get so tangled. As a result, there might not be a solar cycle because twisted tubes of magnetic field might not form and rise through the photosphere to produce sunspots and active regions with prominences and flares. On the other hand, convection might still tangle the magnetic field and produce some activity. Is the magnetic activity that causes sunspots and heats the chromosphere and corona driven mostly by differential rotation, or by convection? Astronomers are not sure, but it seems likely that without differential rotation the Sun would not have a strong magnetic field and resulting high-temperature gas above its photosphere.

This is very speculative, but speculation can be revealing. For example, consider a complementary scenario to the one discussed. ***How do you think the Sun's appearance would differ if it had no convection inside?***

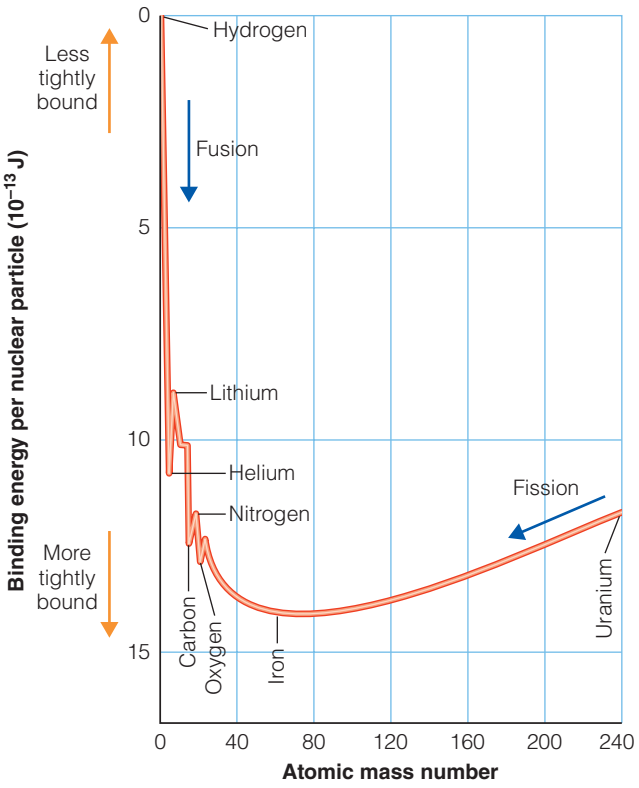
8-3 Nuclear Fusion in the Sun

Like soap bubbles, stars are structures balanced between opposing forces that, if unbalanced, can destroy them. The Sun is a ball of hot gas held together by its own gravity. If it were not for the Sun's gravity, the hot, high-pressure gas in the Sun's interior would explode outward. Likewise, if the Sun were not so hot, its gravity would compress it into a small, dense body.

In this section, you will discover that the Sun is powered by nuclear reactions occurring near its center. The energy released by those reactions keeps the interior hot and the gas totally ionized (meaning, all electrons moving unattached to nuclei). How exactly can the nucleus of an atom yield energy? The answer lies in the force that holds the particles in nuclei together.

8-3a Nuclear Binding Energy

The Sun's luminosity ultimately comes from the forces binding particles inside atomic nuclei. There are only four different ways in which matter affects other matter. These are called the four forces of nature: gravity, the electromagnetic force, the



▲ **Figure 8-15** The orange curve in this graph shows the binding energy per particle, the energy that holds particles inside atomic nuclei. The horizontal axis gives the atomic mass number of each element, the number of protons and neutrons in the nucleus. Both fission and fusion nuclear reactions "move" downward in the diagram (*arrows*), meaning the nuclei produced by a reaction are more tightly bound than the nuclei that went into the reaction, and the reaction resulted in a net release of energy. Iron has the most tightly bound nucleus, so no nuclear reactions can use iron and release energy.

weak nuclear force, and the **strong nuclear force**. The strong nuclear force binds together atomic nuclei, and the weak nuclear force is involved in radioactive decay and other interactions of certain kinds of nuclear particles.

The strong and weak nuclear forces are short-range forces that are effective only within the nuclei of atoms. Nuclear energy originates from the strong force, as nuclear reactions break and re-form the bonds that hold atomic nuclei together. In contrast, the process of burning wood is a chemical reaction that extracts energy by breaking and rearranging chemical bonds among atoms in the wood. The chemical energy released when those bonds are broken and rearranged comes from the electromagnetic force.

There are two types of reactions by which atomic nuclei can release energy. Nuclear power plants on Earth use **nuclear fission** reactions that split massive nuclei into less massive fragments. The isotope of uranium normally used for nuclear fuel contains a total of 235 protons and neutrons. Splitting such a nucleus produces a range of possible fragment nuclei, each containing roughly half as many particles. Because the fragment nuclei are more tightly bound (have lower total potential energy) than the original uranium nucleus, binding energy is released during uranium fission.

Stars make energy by another type of nuclear reaction—**nuclear fusion**—that combines small nuclei into larger, more massive nuclei. The most common reaction inside stars, the one that occurs in the Sun, fuses hydrogen nuclei (single protons) to produce helium nuclei, which contain two protons and two neutrons. Just as with fission, because the nuclei produced by fusion are more tightly bound than the original nuclei, net energy is released.

The curve plotted in **Figure 8-15** shows the nuclear binding energy that holds various atomic nuclei together. If the data point for a given type of nucleus is low in the diagram, the particles in that nucleus are held together tightly. Notice that both fusion and fission reactions involve moving downward in the diagram from less tightly bound toward more tightly bound nuclei. Both types of nuclear reaction produce energy by releasing binding energy of atomic nuclei.

8-3b Hydrogen Fusion

The fusion reaction in the Sun combines four hydrogen nuclei to make one helium nucleus. Because one helium nucleus has 0.7 percent less mass than four hydrogen nuclei, it seems that some mass vanishes in the process. To see this, subtract the mass of a helium nucleus from the mass of four hydrogen nuclei:

$$\begin{array}{rl} 4 \text{ hydrogen nuclei} &= 6.690 \times 10^{-27} \text{ kg} \\ -1 \text{ helium nucleus} &= 6.646 \times 10^{-27} \text{ kg} \\ \hline \text{Difference in mass} &= 0.044 \times 10^{-27} \text{ kg} \end{array}$$

That mass difference, 0.044×10^{-27} kg, does not actually disappear but is converted to energy according to Einstein's famous equation (Eq. 5-3):

$$\begin{aligned} E &= m_0 c^2 \\ &= (0.044 \times 10^{-27} \text{ kg}) \times (3.0 \times 10^8 \text{ m/s})^2 \\ &= 4.0 \times 10^{-12} \text{ J} \end{aligned}$$

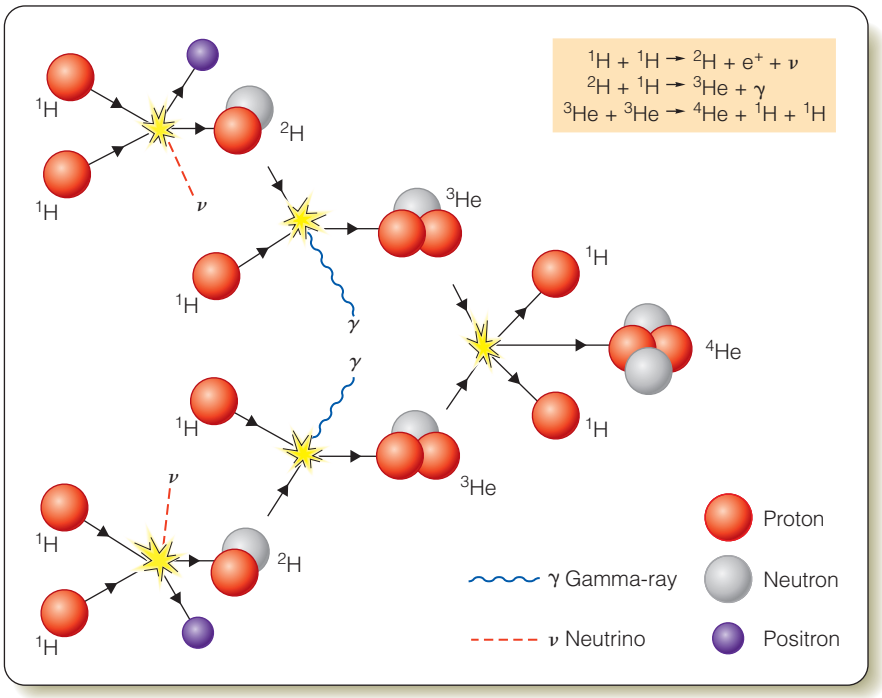
You can symbolize the fusion reactions in the Sun with a simple equation:

$$4\ {}^1\text{H} \longrightarrow {}^4\text{He} + \text{energy}$$

In that equation, the superscripts indicate the number of nucleons (protons plus neutrons) in each of the nuclei. ^{1}H represents a proton—the nucleus of a hydrogen atom—and ^{4}He represents the nucleus of a helium atom.

The actual steps in the process are more complicated than this convenient summary suggests. Instead of waiting for four hydrogen nuclei to collide simultaneously, a highly unlikely event, the process normally proceeds step by step in a series of reactions called the **proton–proton chain** (**Figure 8-16**). The proton–proton chain consists of three nuclear reactions that build a helium nucleus by adding protons one at a time. Those three reactions are:

$$\begin{aligned} {}^1\text{H} + {}^1\text{H} &\longrightarrow {}^2\text{H} + e^+ + \nu \\ {}^2\text{H} + {}^1\text{H} &\longrightarrow {}^3\text{He} + \gamma \\ {}^3\text{He} + {}^3\text{He} &\longrightarrow {}^4\text{He} + {}^1\text{H} + {}^1\text{H} \end{aligned}$$



◀ **Figure 8-16** The proton–proton chain combines four protons (*at far left*) to produce one helium nucleus (*at right.*) Energy is produced mostly as gamma-rays (γ) and as positrons (e^+), which combine with electrons and convert their mass into more gamma-rays. Neutrinos (ν) escape without heating the gas.

In the first reaction, two protons (two hydrogen nuclei) combine. The strong nuclear force binds the protons together, whereas the weak nuclear force causes one of them to transform into a neutron and emit two particles: a **positron**, which is the positively charged version of an electron (e^+); and a **neutrino** (ν), which is a subatomic particle having an extremely low mass and a velocity nearly equal to the velocity of light. The combination of a proton with a neutron forms a heavy hydrogen nucleus called **deuterium** (^2H).

In the second reaction, a deuterium nucleus absorbs another proton and, with the emission of a gamma-ray photon (γ) becomes a lightweight helium nucleus (^3He). Finally, two lightweight helium nuclei combine to form a nucleus of normal helium plus two hydrogen nuclei. Because the last reaction needs two ^3He nuclei, the first and second reactions must occur twice. The net result of this sequence of reactions is the transformation of four hydrogen nuclei into one helium nucleus plus energy.

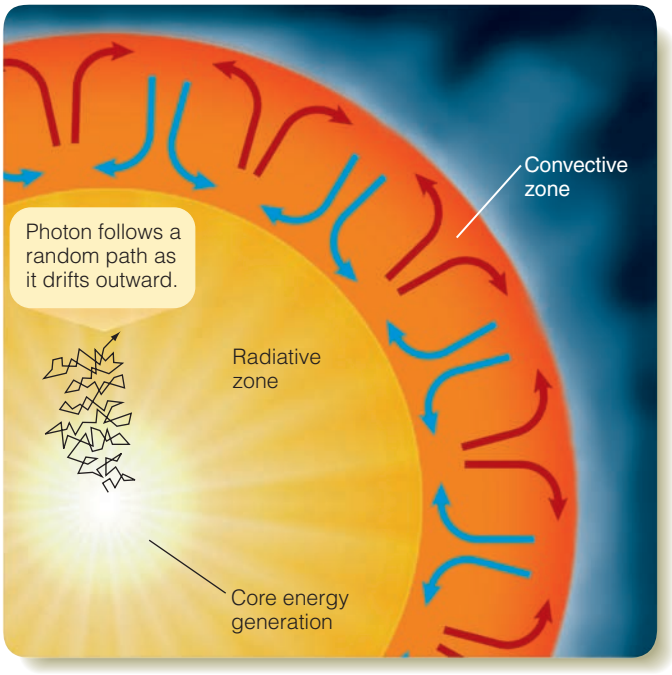
The energy released in the proton–proton chain appears in the form of gamma-rays, positrons, neutrinos, and the energy of motion of all the particles. The gamma-rays are photons that are absorbed by the surrounding gas before they can travel more than a fraction of a millimeter. That heats the gas. The positrons produced in the first reaction combine with free electrons, and both particles vanish, converting their mass into gamma-rays, which are also absorbed and help keep the gas hot. In addition, when fusion produces new nuclei, they fly apart at high speed and collide with other particles. This energy of motion helps raise the temperature of the gas. The neutrinos, on the other hand, don't heat the gas. Neutrinos are particles that almost never interact with other particles. The average neutrino could pass unhindered through a lead wall more than a light-year thick. Consequently, the neutrinos do not warm the gas but race out of the Sun at nearly the speed of light, carrying away approximately 2 percent of the energy produced by the fusion reactions.

Creating one helium nucleus makes only a small amount of energy, not enough to raise a housefly 100 nanometers. Because one reaction produces such a small amount of energy, it is obvious that reactions must occur at a tremendous rate to supply the energy output of a star. The Sun, for example, completes 10^{38} fusion reactions every second, transforming about 4 million tons of matter into energy. It might sound as if the Sun is losing mass at a furious rate, but in its entire estimated 12-billion-year lifetime, the Sun will convert only about 0.1 percent of its mass into energy.

It is a **Common Misconception** that nuclear fusion in the Sun is tremendously powerful. After all, the fusion of a milligram of hydrogen (roughly the mass of a match head) produces as much energy as burning 5 gallons of gasoline. However, at any one time, only a tiny fraction of the hydrogen atoms are fusing into helium, and the nuclear reactions in the Sun are spread through a large volume in its core. Any single gram of matter produces only a little energy. A person of normal mass eating a normal diet produces about 3000 times more heat per gram than the matter in the core of the Sun. Gram for gram, you are a much more efficient heat producer than the Sun. The Sun produces a lot of energy because it contains many grams of matter in its core.

Fusion reactions can occur only when the nuclei of two atoms get very close to each other. Because atomic nuclei carry positive charges, they repel each other with an electrostatic force called the *Coulomb force* (Section 7-1c). Physicists commonly refer to this electrical resistance to nuclear collisions as the **Coulomb barrier**. To overcome this barrier and get close together, atomic nuclei must collide violently. Sufficiently violent collisions are rare unless the gas is very hot, so that the nuclei move at high enough speeds. (Recall that an object's temperature is related to the speed with which its particles move.) Even so, the fusion of two protons is a highly unlikely process. If you could follow a single proton in the Sun's core, you would see it encountering and bouncing off other protons millions of times a second, but you would have to follow it around for many billions of years before it would have a 50/50 chance of penetrating the Coulomb barrier and combining with another proton.

Because of the dependence of nuclear reactions on particle collisions, the reactions in the Sun take place only near its center, where the gas is hot and dense. A high temperature ensures that collisions between nuclei are violent, and a high density ensures that there are enough collisions, and thus enough reactions per second, to make energy at the Sun's rate. The proton–proton chain requires temperatures above about 4 million K.



▲ **Figure 8-17** A cross section of the Sun's interior. Near the center, nuclear fusion reactions sustain high temperatures. Energy flows outward through the radiative zone as photons that gradually make their way to the surface as they are randomly deflected over and over by collisions with electrons. In cooler, more opaque outer layers the energy is carried by rising convection currents of hot gas (*red arrows*) and sinking currents of cooler gas (*blue arrows*).

8-3c Energy Transport in the Sun

Now you are ready to follow the energy from the core of the Sun to the surface. You will learn in a later chapter that astronomers have computed models indicating that the temperature at the center of the Sun must be about 16 million K for the Sun to be stable. Compared with that, the Sun's surface is very cool, only about 5800 K. Heat always moves from hot regions to cool regions, so energy must flow from the Sun's high temperature core outward to the cooler surface where it is radiated into space.

Because the core is so hot, the photons there are gamma-rays. Each time a gamma-ray encounters one of the matter particles—electrons or nuclei—it is deflected, or scattered, in a random direction, and, as it bounces around, it slowly drifts outward toward the surface while being converted into several photons of lower energy. The net outward motion of energy in the inner parts of the Sun takes the form of radiation, so astronomers refer to that region as the **radiative zone**.

Energy originally produced in the core of the Sun and traveling outward as radiation eventually reaches the outer layers of the Sun where the gas is cool enough that it is not completely ionized. Partially ionized gas is much less transparent to radiation than is completely ionized gas. So, at that point, the energy flowing toward the Sun's surface backs up like water behind a dam, and the gas begins to churn in convection currents. Hot blobs of gas rise, and cool blobs sink. In this region, known as the **convective zone**, the energy is carried outward not as photons but as circulating gas (Figure 8-17). Rising hot gas carries energy outward, but sinking cool gas is a necessary part of the cycle. The result is net transport of energy continuing outward. Previously in this chapter, you learned about granulation and supergranulation features observed on the Sun's photosphere; those are the visible effects of energy arriving at the Sun's surface from its interior by convection.

It can take millions of years for the energy that began in the form of a single gamma-ray produced in the center of the Sun to work its way outward first as radiation and then by convection. When that energy finally reaches the photosphere, it is radiated into space as about 2000 photons of visible light.

It is time to ask the critical question that lies at the heart of science. What is the evidence to support this theoretical explanation of how the Sun shines?

8-3d Counting Solar Neutrinos

Nuclear reactions in the Sun's core produce floods of neutrinos that rush out of the Sun and off into space. More than 10^{14} (100 trillion) solar neutrinos flow through your body every second, but you never feel them because you are almost perfectly transparent to neutrinos. If you could detect these neutrinos, you could probe the Sun's interior. You can't focus neutrinos with a lens or mirror, and they zip right through detectors used

HOW DO WE KNOW? 8-2

Scientific Confidence

How can scientists be certain of something?

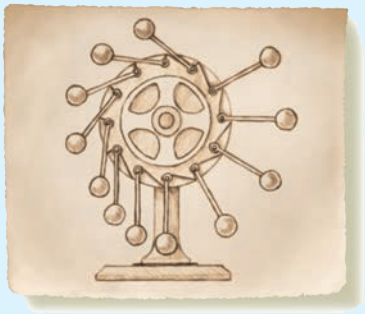
Sometimes scientists stick so firmly to their ideas in the face of contradictory claims that it almost sounds as if they are merely stubbornly refusing to consider alternatives. To understand what's actually going on, you might consider the perpetual motion machine, which is a device that supposedly runs continuously with no source of energy. If you could invent a real perpetual motion machine, you could make cars that would run without any fuel. That's good mileage.

For centuries many people have claimed to have invented perpetual motion machines, and for just as long scientists have been dismissing these claims as impossible. The problem with a perpetual motion machine is that it violates the law of conservation of energy, and scientists are not willing to accept that the law could be wrong. In fact, the Royal Academy of Sciences in Paris was so sure that a perpetual motion machine is impossible, and so tired of debunking hoaxes, that in 1775 they issued a formal statement refusing to deal with them. The U.S. Patent Office policy is that it won't even consider starting the patent process for one without seeing a working model first.

Why do scientists seem so stubborn and close-minded on this issue? Why isn't one person's belief in perpetual motion just as valid as another person's belief in the law of conservation of energy? In fact, the two positions are not equally valid. The confidence physicists have in that law is not a belief or even an opinion; it is an understanding founded on the fact that the law has been tested uncountable times and has never failed. In contrast, no one has ever successfully demonstrated a perpetual motion machine. The law of conservation of energy is a fundamental truth about nature and can be used to understand what is possible and what is impossible.

When the first observations of solar neutrinos detected fewer than were predicted, some scientists speculated that astronomers misunderstood how the Sun makes its energy or that they misunderstood the internal structure of the Sun. But astronomers stubbornly refused to reject their model because the nuclear physics of the proton–proton chain is well understood, and models of the Sun's structure have been tested successfully by other measurements many times. The confidence astronomers felt in their understanding of the Sun was an example of scientific certainty, and that confidence in basic natural laws prevented them from abandoning decades of work in the face of a single contradictory observation.

What seems to be stubbornness among scientists is really their confidence in basic principles that have been tested over and over. Those principles are the keel that keeps the ship of science from rocking before every little breeze. Without even looking at that perpetual motion machine, your physicist friends can warn you not to invest in it.



For centuries, people have tried to design a perpetual motion machine, but not a single one has ever worked. Scientists understand why.

to count other atomic particles, but neutrinos of certain energies can trigger the radioactive decay of some atoms. That gives astronomers a way to detect neutrinos.

In the 1960s, chemist Raymond Davis Jr. created a device that could count neutrinos with energies produced by hydrogen fusion in the Sun. He buried a 100,000-gallon tank of cleaning fluid (perchloroethylene, C_2Cl_4) in the bottom of a South Dakota gold mine where other types of cosmic rays could not reach it and invented a way to count individual argon atoms that were produced by neutrinos colliding with chlorine atoms in the tank.

The Davis neutrino experiment created a huge controversy. It was expected to detect one neutrino a day, but it actually counted one-third as many: Only one solar neutrino was captured in that 100,000-gallon tank every three days. Were scientists wrong about nuclear fusion in the Sun? Did they misunderstand how neutrinos behave? Was the detector not working properly? Because astronomers had reason for confidence in their understanding of the solar interior, they didn't immediately abandon their hypotheses (**How Do We Know? 8-2**).

It took more than 30 years, but eventually physicists were able to build better and different neutrino detectors (**Figure 8-18**). They discovered that neutrinos change back and forth among three different types, which physicists call

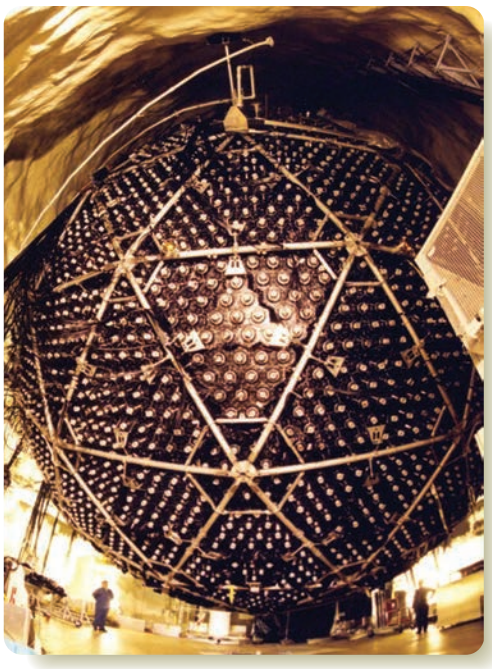


Photo courtesy of SNO

▲ **Figure 8-18** The Sudbury Neutrino Observatory is a 12-meter-diameter globe containing water rich in deuterium (heavy hydrogen) in place of ordinary hydrogen. Buried 2100 m (6800 ft) down in an Ontario mine, it can detect all three flavors of neutrinos and confirms that neutrinos oscillate.

"flavors." Nuclear reactions in the Sun produce just one flavor, and the Davis experiment was designed to detect (taste!) only that flavor. But during the 8-minute journey from the Sun's core to Earth, the neutrinos changed flavor so many times that they were distributed evenly among the three different flavors by the time they arrived at Earth. That's why the Davis experiment detected only one-third of the number originally predicted. Also, in 2014, neutrinos from the first step of the proton–proton chain were detected at the predicted rate; these neutrinos are more numerous but have lower energy and thus are more difficult to detect than the ones measured by the Davis experiment. Models of nuclear fusion in the Sun are now fully confirmed by the new observations plus improved theoretical understanding of neutrino properties.

The center of the Sun seems forever beyond human experience, but counting solar neutrinos provides evidence to confirm the theories. The Sun makes its energy through nuclear fusion.

Practicing SCIENCE

Why does nuclear fusion require that the gas be very hot? Is there any alternative? Answering these questions requires scientists to rehearse what is known about the basic physics of atoms and thermal energy. Occasional reconsidering and questioning of basic principles are important parts of practicing science.

Inside a star, the gas is so hot it is ionized, which means the electrons have been stripped off the atoms, leaving bare, positively charged nuclei. In the case of hydrogen, the nuclei are single protons. These atomic nuclei repel each other because of their positive charges, so they must collide with each other at high speed if they are to overcome that repulsion and get close enough together to fuse. If the atoms in a gas are moving rapidly, then the gas must have a high temperature, so nuclear fusion requires that the gas be very hot. If the gas is cooler than about 4 million K, hydrogen can't fuse because the protons don't collide violently enough to overcome the repulsion of their positive charges. In spite of rumors to the contrary, there does not seem to be any "shortcut" allowing fusion to happen at much lower temperatures.

It is easy to see why nuclear fusion in the Sun requires a high temperature, but now consider a related question about basic physical laws: ***Why is fusion helped by high density?***

WHAT ARE WE?

Sunlight

We live very close to a star and depend on it for survival. All of our food comes from sunlight that was captured by plants on land or in the oceans. We either eat those plants directly or eat the animals that feed on those plants. Whether you had salad, seafood, or a cheeseburger for supper last night, you dined on sunlight, thanks to photosynthesis.

Almost all of the energy that powers human civilization came from the Sun through photosynthesis by ancient plants that were buried and converted to coal, oil, and natural gas. New technology is making energy from plant products like corn, soybeans, and sugar. It is all stored sunlight. Windmills generate electrical power, but the wind blows because of heat from the Sun. Photocells make electricity directly from sunlight. Even our bodies have adapted to use sunlight to help manufacture vitamin D.

Our planet is warmed by the Sun; without that warmth the oceans would be ice, and the atmosphere would be a coating of frost. Books often refer to the Sun as "our Sun" or "our star." It is ours in the sense that we are creatures of its light.

Study and Review

SUMMARY

8-1 The Solar Photosphere and Atmosphere

- The Sun's visible surface, the **photosphere**, is the layer in the Sun from which visible photons most easily escape. The solar atmosphere can be considered to consist of the photosphere plus two layers of hotter, lower-density gas above the photosphere: the **chromosphere** and **corona**.
- **Granulation** of the photosphere is produced by **convection** currents of hot gas rising from below, which cool and sink below the visible surface. **Supergranules** appear to be caused by larger convection currents rising from deeper than those currents producing the granules. The edge, or **limb**, of the solar disk is dimmer than the center. This **limb darkening** effect is evidence that the temperature in the solar photosphere increases with depth.
- The chromosphere is a thin, hot layer of gas just above the photosphere, most easily visible during total solar eclipses. **Filtergrams** of the chromosphere reveal **spicules**, which are flamelike structures that extend upward into the lower corona.
- The corona can be imaged using a **coronagraph**. The current hypothesis is that the corona's high temperature is maintained by energy transported via motions of the magnetic field extending up through the photosphere—the **magnetic carpet**—and by magnetic waves coming from below the photosphere. Parts of the corona give rise to the **solar wind**, which extends to the **heliopause**, marking an outer boundary of the Solar System.

- The strength of spectral lines in the Sun's spectrum depends partly on the temperature of its atmosphere. Some elements that are abundant have weak spectral lines, and some that are not very abundant have strong spectral lines.
- Solar astronomers can study the motion, density, and temperature of gases inside the Sun by analyzing the way the solar photosphere oscillates. Known as **helioseismology**, this field of study requires large amounts of data and extensive computer analysis.

8-2 Solar Activity

- The Sun's light and infrared radiation can burn your eyes, so you must take great care in observing it. **Sunspots** come and go on the Sun, but only rarely are they large enough to be visible to the unaided eye.
- Sunspots seem dark to our eyes because they are slightly cooler than the rest of the photosphere. The average sunspot is about twice the size of Earth. Sunspots appear for a month or so and then fade away, and the number and location of spots on the Sun vary with an 11-year sunspot cycle.
- Early in a sunspot cycle, spots appear farther from the Sun's equator, and later in the cycle they appear closer to the equator. This repetitive pattern of sunspot locations over time is shown in the **Maunder butterfly diagram**.
- Astronomers can use the **Zeeman effect** to measure magnetic fields on the Sun. The average sunspot contains magnetic fields a few thousand times stronger than Earth's field. This increase in strength is part of the evidence that magnetic fields are involved in the sunspot cycle. Two sunspot cycles occur for every one solar magnetic cycle, which is 22 years long.
- The sunspot cycle does not always have the same pattern. For example, during the **Maunder minimum** from 1645 to 1715, the number of sunspots was significantly fewer, solar activity was very low, and Earth's climate was slightly colder.
- Sunspots are the visible consequences of **active regions** where the Sun's magnetic field is strong. Arches of magnetic field can produce sunspots where the field passes through the photosphere.
- The Sun's magnetic field is produced by the **dynamo effect** operating at the base of the zone of convection currents. Alternate sunspot cycles have reversed magnetic **polarity**, which has been explained by the **Babcock model**, in which the Sun's **differential rotation** and convection currents tangle the magnetic field. As the Sun's magnetic field becomes strongly tangled, **meridional flow** currents help reorder the magnetic field into a simpler but reversed field, and the sunspot cycle starts over.
- Arches of magnetic field are visible as **prominences** in the chromosphere and corona. Seen from above in filtergrams, prominences are visible as dark **filaments** silhouetted against the bright chromosphere.
- Magnetic field **reconnection events** can produce powerful **flares**, which are sudden eruptions of X-ray, ultraviolet, and visible radiation plus high-energy atomic particles. Flares are important because they can have dramatic effects on Earth, such as communications blackouts.
- The solar wind originates in regions on the solar surface called **coronal holes**, where the Sun's magnetic field leads out into space and does not loop back to the Sun. **Coronal mass ejections (CMEs)** occur when magnetic fields on the surface of the Sun eject bursts of ionized gas that flow outward in the solar wind. If they strike Earth, such bursts can produce **auroras** and other phenomena.
- Other stars are too far away to observe visible star spots, which are the equivalents of sunspots. However, some stars vary in brightness in ways that reveal they do indeed have spots on their surfaces. Spectroscopic observations reveal that many other stars have spots and magnetic fields that follow long-term cycles like the Sun's.
- The **solar constant** is a measure of all the electromagnetic radiation reaching Earth from the Sun; its value is about 1370 J/m^2/s. Sensitive measurements show that the solar constant is not truly constant but varies during a sunspot cycle and perhaps on longer time scales.

8-3 Nuclear Fusion in the Sun

- The **strong nuclear force** binds atomic nuclei together. The **weak nuclear force** is involved in radioactive decay and other interactions of certain kinds of nuclear particles. Nuclear reactors on Earth generate energy through **nuclear fission**, in which large nuclei such as uranium break into smaller fragments and generate energy in the process. The Sun generates its energy through **nuclear fusion**, in which small nuclei such as hydrogen fuse to form larger nuclei, such as helium, generating energy in the process.
- The fusion of hydrogen into helium in the Sun proceeds in three steps known as the **proton–proton chain**. The first step in the chain combines two hydrogen nuclei to produce a heavy hydrogen nucleus called **deuterium**. The second step forms light helium, and the third step combines the light helium nuclei to form normal helium. During the process, **positrons**, **neutrinos**, and gamma-rays are formed and energy is released as the particles fly away. Fusion can occur only in the core of the Sun where temperatures are pressures are high enough to overcome the **Coulomb barrier** and fuse together.
- Energy is transported from the core of the Sun's by photons traveling through the **radiative zone** to the base of the **convective zone**. Then energy is transported to the photosphere by rising currents of hot gas and sinking currents of cooler gas through this convective zone.
- Neutrinos escape from the Sun's core at nearly the speed of light, carrying away about 2 percent of the energy produced by fusion. Experiments detecting solar neutrinos on Earth, confirm the hypothesis that much of the Sun's energy comes from the proton–proton chain.

REVIEW QUESTIONS

1. Why can't you see deeper into the Sun than the photosphere?
2. What color is the photosphere as viewed from the ground on a clear, cloudless day when the Sun is highest overhead? When the Sun has sunk to just above the ocean's horizon? When the Sun has sunk to half below the ocean's horizon? Does the photosphere really change colors during this sunset? Why or why not?
3. You stayed a little too long outside in the sunshine, and now you have a sunburn. From which atmospheric layer of the Sun did the photons originate that resulted in your sunburn? How do you know?
4. The average temperature of the photosphere is 5780 K. What color is the maximum intensity of a 5780 K blackbody? Is this the color we normally associate with the photosphere? Why or why not? (*Hint:* Refer to Figure 7-7 and Section 8-1.)
5. What information can we obtain from the Sun's absorption line spectrum shown in Figure 7-8a?
6. Which atmospheric layer is associated with the Sun's continuous spectrum? With its absorption spectrum? With its emission spectrum?
7. What evidence can you give that granulation is caused by convection?
8. How are granules and supergranules alike? How are they different?
9. How can astronomers detect structure in the chromosphere?
10. What evidence can you give that the corona has a very high temperature?
11. What heats the chromosphere and corona to maintain such high temperatures?

12. Why does hydrogen, which is abundant in the Sun's atmosphere, have relatively weak spectral lines, whereas calcium, which is not abundant, has very strong spectral lines?
13. What is the shape of the heliopause? What determines that shape?
14. How are astronomers able to explore the layers of the Sun below the photosphere?
15. Energy can be transported by convection, conduction, and radiation. Which of these is (or are) associated with the interior of the Sun?
16. What evidence can you give that sunspots are magnetic?
17. When in the cycle does the maximum number of sunspots occur? Is it at the beginning, in the middle, or at the end of a sunspot cycle? Is this time in the cycle reflected in the Maunder butterfly diagram? Why or why not?
18. How does the Babcock model explain the sunspot cycle?
19. How is meridional flow related to the Sun's magnetic dynamo, the sunspot cycle, and the Maunder butterfly diagram?
20. *Meridional* is derived from *meridian*. Look up the definition of meridian; what is the direction of solar meridional flow? How is it detected?
21. What does the spectrum of a prominence reveal? What does its shape reveal?
22. How do prominences affect Earth?
23. How can solar flares affect Earth?
24. Which has a more tightly bound nucleus, uranium or helium? How do you know?
25. Why does nuclear fusion require high temperatures and high densities?
26. Why does nuclear fusion in the Sun occur only near the center?
27. Four protons are combined in the proton–proton chain to make one helium nucleus. But, a helium nucleus only contains two protons. What has happened to two of the protons from the first step of the chain?
28. Give an example of a charged subatomic particle and a neutral subatomic particle discussed in this chapter.
29. Examine the elemental abundances for the solar atmosphere in Table 8-1. Then, look at Figure 8-17, a cross-section of the Sun showing the core where fusion occurs and the outer zone containing convective currents. Do you think the atmospheric composition has changed since the Sun formed? Why or why not?
30. How can astronomers detect neutrinos from the Sun?
31. How did neutrino oscillation affect the detection of solar neutrinos by the Davis experiment?
32. **How Do We Know?** How do confirmation and consolidation extend scientific understanding?
33. **How Do We Know?** What does it mean when scientists say they are certain? What does scientific certainty really mean?

ACTIVE INQUIRY QUESTIONS

1. Explain why the presence of spectral lines of a given element in the solar spectrum tells you that element is present in the Sun, but the absence of the lines would not necessarily mean the element is absent from the Sun.
2. What energy sources on Earth *cannot* be thought of as stored sunlight? Why?
3. What would the spectrum of an auroral display look like? Why?
4. What observations would you make if you were ordered to set up a system that could warn orbiting astronauts of dangerous solar flares? (Such a warning system actually exists.)

PROBLEMS

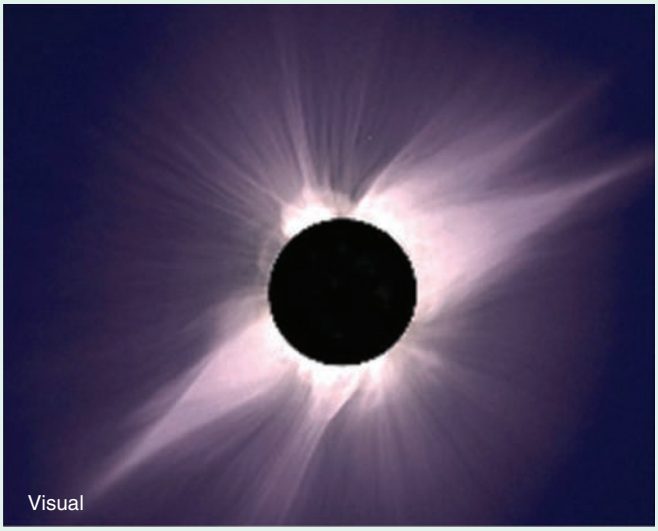
1. The radius of the Sun is 0.7 million km. What percentage of the radius is taken up by the chromosphere? (*Hint:* Refer to Figure 8-1.)
2. What fraction of the Sun's interior is the core? The radiative zone? The convective zone? Which layer occupies most of the volume of the Sun's interior? (*Hint:* Refer to Figure 8-17.)
3. The smallest detail visible with ground-based solar telescopes is about 1 arc second. How large a region does this represent on the Sun? (*Hint:* Use the small-angle formula, Eq. 3-1.) (*Note:* The average distance to the Sun can be found in Celestial Profile: The Sun.)
4. What is the angular diameter of a star the same size as the Sun located 5 light-years from Earth? Is the *Hubble Space Telescope* able to detect detail on the surface of such a star? (*Hint:* Use the small-angle formula, Eq. 3-1.)
5. If a sunspot has a temperature of 4200 K and the sunspot can be considered a blackbody, what is the wavelength of maximum intensity in nm units and what color is associated with this wavelength? Is this the color we see the sunspot as from Earth? Why or why not? (*Hint:* Refer to Wien's law, Eq. 7-2.)
6. How many watts of radiation does a 1-meter-square region of the Sun's photosphere emit, at a temperature of 5780 K? How much would the wattage increase if the temperature were twice as much, 11,600 K? (*Hint:* Use the Stefan-Boltzmann law, Eq. 7-1.)
7. If a sunspot has a temperature of 4200 K and the average solar photosphere has a temperature of 5780 K, how much more energy is emitted in 1 second from a square meter of the photosphere compared to a square meter of the sunspot? (*Hint:* Use the Stefan-Boltzmann law, Eq. 7-1.)
8. The radial velocity of a granule's center is found to be –0.4 km/s. If the observed spectral line is Balmer H-alpha at a laboratory wavelength of 656.300 nm, at what wavelength is the line observed? Is that a blueshift or a redshift? Does that mean the gas is rising, sinking, or moving laterally across the line of sight? (*Hint:* Use the Doppler formula, Eq. 7-3.)
9. Gusts of the solar wind travel as fast as 1000 km/s. How many days would the solar wind take to reach Earth at this speed? (*Note:* The average distance to the Sun can be found in Celestial Profile: The Sun.)
10. If the Sun rotates once every 24.5 days at the equator, once every 27.8 days at latitude 45 degrees, and once every 34.3 days at the poles, what is the average number of days the Sun takes to rotate once based on these three numbers?
11. How much energy is produced when the Sun converts 1 kg of mass into energy?
12. How much energy is produced when the Sun converts 1 kg of hydrogen into helium? (*Hint:* How does this problem differ from Problem 11?)
13. A 1-megaton nuclear weapon produces about 4×10^{15} J of energy. How much mass must vanish when a 1-megaton weapon explodes?
14. A solar flare can release 10^{25} J. How many megatons of TNT would be equivalent? (See Problem 13 for conversion between megatons and joules.)
15. The United States consumes about 2.5×10^{19} J of energy in all forms in a year. How many years could you run the United States on the energy released by the solar flare in Problem 14?
16. Use the luminosity of the Sun (the total amount of energy the Sun emits each second) to calculate how much mass the Sun converts into energy each second.

SENSE OF PROPORTION

1. The diameter of the Sun is 109 times the diameter of Earth. How many times larger is the Sun's surface area than Earth's surface area? How many times larger is the Sun's volume than Earth's volume?
2. Arrange the following objects in order of increasing size: a large sunspot group; Earth; a granule; a large prominence

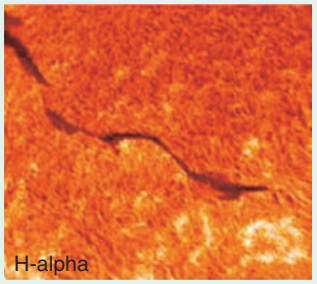
LEARNING TO LOOK

1. Whenever there is a total solar eclipse, you can see something like the image shown here. Explain why the shape and extent of the glowing gases are observed to be different for each eclipse.

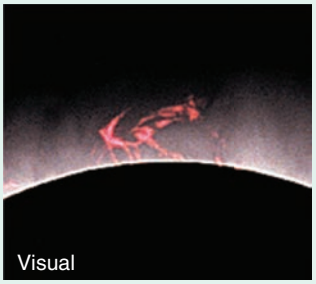


NOAO/AURA/NSF

2. Look at Figure 8-3. To what height and which atmospheric layer does the red color correspond? How about the yellow color?
3. Refer to Figure 7-8a and Figure 8-3. What kinds of spectra are shown in Figure 7-8a? What atmospheric layers of the Sun are associated with these kind of spectra? At what height in km is the gas that generated these kind of spectra?
4. Refer to labels C, F, and G in the solar spectrum shown in Figure 7-8a. What are the wavelengths of those lines? Which hydrogen spectral line series are these lines in? (*Hint:* Refer to Chapter 7 Concept Art: Atomic Spectra.)
5. The two images here show two solar phenomena. What are they, and how are they related? How do they differ?

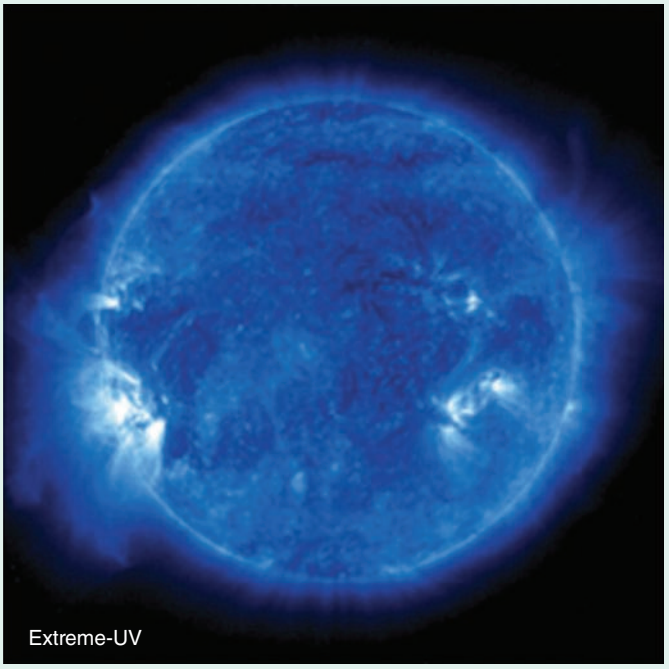


NOAO/AURA/NSF



NOAO/AURA/NSF

6. This image of the Sun was recorded in the extreme ultraviolet by the *SOHO* spacecraft. Explain the features you see.



ESA/NASA/*SOHO*

9 The Family of Stars

Y. Beletsky/ESO

▲ Star cluster NGC 3766, located about 5500 ly away in the constellation Centaurus. Many luminous blue main-sequence stars and two red supergiants can be seen.

Measurement is fundamental to science, but making measurements of distant celestial objects is often difficult. To discover the properties of stars, astronomers have used telescopes, photometers, cameras, and spectrographs in clever ways to learn the secrets hidden in starlight. The result is a family portrait of the stars.

In this chapter, you will find answers to five important questions about stars:

- How far away are the stars?
- How much energy do stars make?
- How do spectra of stars allow you to determine their temperatures?
- How big are stars?
- How much mass do stars contain?

Now you are leaving our Sun behind and beginning your study of the billions of stars that dot the sky. In a sense, stars are the basic building blocks of the Universe. If you hope to understand what the Universe is, what our Sun is, what our Earth is, and what we are, you need to understand stars.

Once you know how to find the basic properties of stars, you will be ready to trace the history of the stars from birth to death, a story that begins in the next chapter.

> [Love] is the star to every wandering bark,
> Whose worth's unknown, although his height be taken.
>
> WILLIAM SHAKESPEARE, *SONNET 116*

SHAKESPEARE COMPARED LOVE to a star that can be seen easily and even used to guide a ship, but whose real nature is utterly unknown. Shakespeare was born during the same year as Galileo and, like everyone else at the time, had no idea what stars actually are. To understand the history of the Universe, the origin of Earth, and your place in the cosmos, you can start by discovering what people in Shakespeare's time did not know—the real nature of the stars.

But you run into a problem right away. It is very difficult to find out the real properties of stars. When you look at a star even through a large telescope, you see only a point of light. True understanding of stars requires careful analysis of starlight. This chapter concentrates on five goals: knowing how far away stars are, how much energy they emit, what their surface temperatures are, how large they are, and how much mass they contain. By the time you finish this chapter, you will know how our star, the Sun, fits into the family of stars.

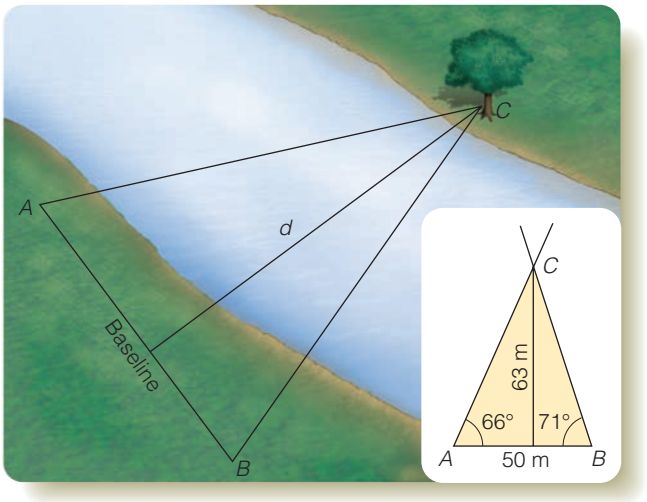
9-1 Star Distances

To find out almost anything else about a star, you first need to know how far away it is. Knowing distances is crucial, but it is also the most difficult measurement in astronomy. Astronomers have found a number of ways to estimate the distance to far-away objects, but each of those ways ultimately depends on a direct geometrical method that is much like the method surveyors use to measure distances to objects they cannot reach directly.

9-1a The Surveyor's Triangulation Method

To measure the distance to a landmark such as a tree on the other side of a river, a team of surveyors begins by driving two stakes into the ground a known distance apart (Figure 9-1). The distance between the stakes is the baseline of the measurement. Using surveying instruments, they sight the tree on the far bank from the two ends of the baseline and measure the two angles on their side of the river. This establishes a large triangle with corners marked by the two stakes and the tree.

Now that the surveyors know two angles of this large triangle and the length of the side between the angles, they can find the distance across the river by simple trigonometry. For example, if the baseline is 50 m long and the angles are 66 and 71 degrees, the distance from the baseline to the tree across the river must be 63 m.



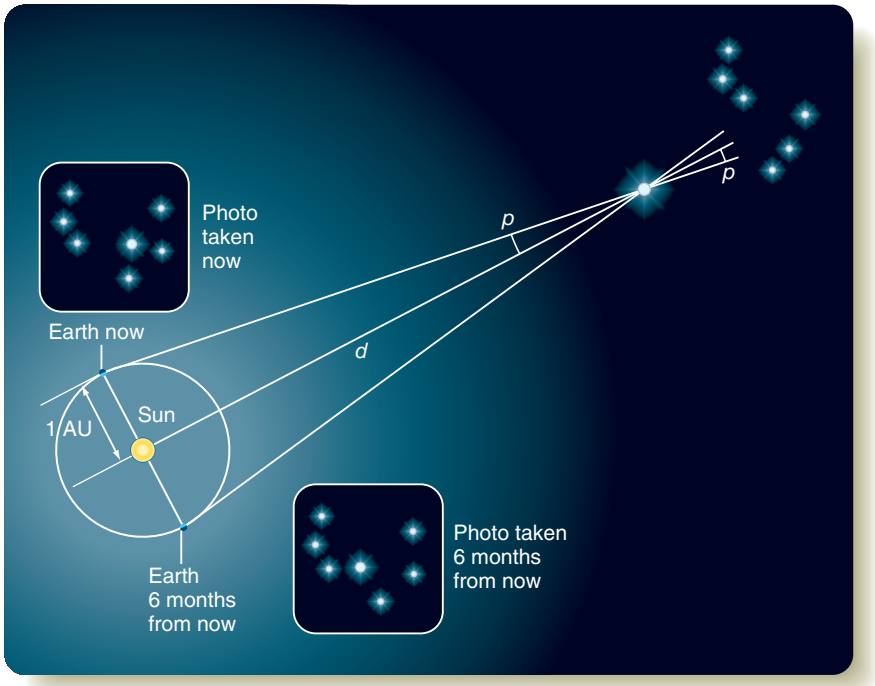
▲ **Figure 9-1** You can find the distance *d* across a river by measuring the length of the baseline and the angles *A* and *B* and then constructing a scale drawing of the triangle or using trigonometry to calculate it.

The more distant an object is, the longer the baseline needed to measure its distance accurately. You could use a baseline 50 m long to find the distance across a river, but to measure the distance to a mountain on the horizon, you might need a baseline 1000 m long. Measurement of greater distances requires longer baselines.

9-1b The Astronomer's Triangulation Method

To find the distances to stars, a very long baseline, the size of Earth's orbit, is used. If you take a photograph of a nearby star and then wait 6 months, Earth will have moved halfway around its orbit. You can then take another photograph of the star from the other side of Earth's orbit. In this example, your baseline equals the diameter of Earth's orbit—2 astronomical units (AU)—and lines of sight to the star from the two observing positions outline a long, thin triangle (Figure 9-2).

When you compare the two photographs taken from different locations around Earth's orbit, you will discover that the nearby star is not seen in exactly the same place relative to background stars. This shift in the position of the star is called **parallax**, which is the apparent change in the position of an object as a result of a change in the location of the observer. In Figure 1a of the Chapter 4 Concept Art: An Ancient Model of the Universe, you saw an everyday example. A thumb, held at arm's length, appears to shift position against a distant background when viewed first with one eye and then with the other. In that case, the baseline is the distance between the observer's eyes, and the parallax is the angle through which the thumb appears to move when the observer changes his or



▲ **Figure 9-2** You can measure the parallax of a nearby star by photographing it from two points around Earth's orbit. For example, you might photograph it now and again in 6 months, when Earth's position has shifted by 2 AU.

her viewing eye. The farther away the observer holds his or her thumb, the smaller will be the thumb's parallax as viewed from opposite ends of the same baseline, the distance between the observer's eyes.

9-1c Parallax and Distance

Because the stars are so distant, their parallaxes are very small angles, usually expressed in arc seconds. Astronomers' conventional definition of **stellar parallax *(p)*** is the shift of the star observed across a 1 AU baseline. That is equal to the radius, rather than the diameter, of Earth's orbit, but this is just a detail. Astronomers measure the parallax, and surveyors measure the angles at the ends of the baseline, but both measurements reveal the same thing: the shape and size of the triangle with a known baseline and thus the distance to the object in question.

To find the distance to a star from its measured parallax, astronomers use the same calculation you have already seen in the small-angle formula (Eq. 3-1). Imagine instead that you could observe our Solar System from the star. Figure 9-2 shows that the maximum angular separation you would measure between the Sun and Earth equals the star's parallax, p. Recall that the small-angle formula relates an object's angular diameter, its linear diameter, and its distance:

$$\frac{\text{angular diameter (in arc seconds)}}{2.06 \times 10^5} = \frac{\text{linear diameter}}{\text{distance}} \qquad \text{(Eq. 3-1)}$$

(Recall from Chapter 3 that the constant 2.06×10^5 is a conversion factor, the number of arc seconds in a radian.)

For the stellar parallax case in Figure 9-2, the angular diameter is the parallax angle p and the linear diameter—the base of the triangle—is 1 AU. Then the small-angle formula, rearranged slightly, tells you that the distance, d, to the star in AU is equal to 2.06×10^5 divided by the parallax, p, in arc seconds:

$$d(\text{AU}) = \frac{2.06 \times 10^5}{p\,(\text{arc seconds})}$$

Because the parallaxes of even the nearest stars are less than 1 arc second, the distances in AU are inconveniently large numbers. To keep the numbers manageable, astronomers have defined the **parsec (pc)** as the primary unit of stellar distances. One parsec equals 2.06×10^5 AU, so the parallax equation becomes simpler:

KEY EQUATION 9-1 Parallax and Distance

$$d(\text{pc}) = \frac{1}{p\,(\text{arc seconds})}$$

Thus, 1 parsec is the distance to an object with a parallax of 1 arc second measured with a 1 AU baseline.

Example: The star Altair has a parallax of 0.195 arc second. How far away is it? ***Solution:*** The distance in parsecs equals 1 divided by 0.195, or 5.13 pc:

$$d(\text{pc}) = \frac{1}{0.195} = 5.13 \text{ pc}$$

One parsec equals 3.26 light-years, so Altair is 16.7 ly away.

The parsec unit is used routinely by astronomers because it simplifies calculation of distance from observed parallax shifts. However, there are instances when units of light-years (ly) are also convenient. In the chapters that follow this one, parsecs or light-years are used as appropriate.

Measuring stellar parallaxes is very difficult because they are such small angles. The visible star nearest the Sun is Famous Star Alpha Centauri. It has a parallax of only 0.747 arc second, and more distant stars have even smaller parallaxes. To see how small these angles are, hold a piece of paper edgewise at arm's length. The thickness of the paper covers an angle of about 30 arc seconds.

The blurring caused by Earth's atmosphere smears star images, making them at least 0.5 to 1 arc second in diameter at even the best observatory sites, so parallaxes are difficult to measure from Earth's surface. Even when astronomers average together many observations, they cannot measure parallax from an Earth observatory with an uncertainty smaller than about 0.002 arc second. Therefore, if

you measure a parallax of 0.02 arc second from Earth, corresponding to a distance of 50 pc (about 160 ly), the uncertainty is about 10 percent, which is about the largest uncertainty in a parallax measurement that astronomers are willing to tolerate. Starting with the first stellar parallax measured in 1838, astronomers using ground-based telescopes have been able to measure accurate parallaxes for only about 10,000 stars closer than about 50 pc.

In 1989, the European Space Agency launched the satellite *Hipparcos* to measure stellar parallaxes from orbit, above the blurring effects of Earth's atmosphere. The satellite observed for four years, and the data were used to produce two parallax catalogs: one listing 120,000 stars with parallaxes 20 times more precise than ground-based measurements, and the other listing more than a million stars with parallaxes about as accurate as ground-based values. *Hipparcos* data have given astronomers new insights into the nature of stars.

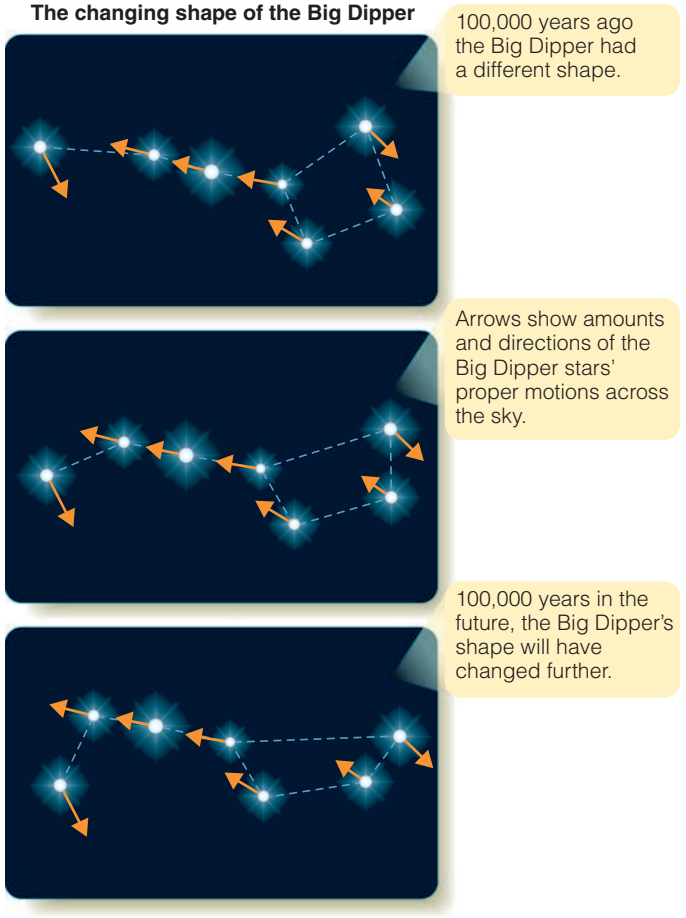
The European Space Agency launched the *Gaia* spacecraft in 2013. It is designed to measure the parallaxes of a billion stars as faint as apparent visual magnitude +20. This will allow the first real three-dimensional map of our Galaxy; an initial catalog containing 2 million parallaxes was released in 2016.

9-1d Proper Motion

All the stars in the sky, including the Sun, are moving along orbits around the center of our Galaxy. That motion isn't obvious to the eye during a human lifetime, but over centuries it can significantly distort the shape of constellations (**Figure 9-3**).

If you make images of a small area of the sky a number of times over the course of several years, precise measurements will reveal not only cyclical parallax motions but also that some of the stars have moved continuously against the background of more distant stars. This motion, called the **proper motion** of the stars, is usually expressed in units of arc seconds per year. As examples, consider two stars from the Famous Stars list. Vega, a bright blue-white star in the summer sky, has a proper motion of 0.350 arc second per year. Rigel, a bright blue-white star in the winter sky, has a proper motion of only 0.002 arc second per year. The two stars are nearly the same brightness in the sky and have about the same color and temperature, but the proper motion of Rigel is less than 1/100 that of Vega. What does this mean?

A star might have a small proper motion if it is moving almost directly toward or away from you; then its position on the sky would change only slowly. That is unlikely, but it does happen. Another reason a star might have a small proper motion is that it could be quite far away from you. Then, even if the star were moving rapidly through space, it would have a small proper motion. That explains why Rigel, at a distance of 260 pc (determined by *Hipparcos* parallax measurements), has a smaller proper motion, and Vega, at a distance of only 7.7 pc, has a larger proper motion. Although Rigel is 34 times farther away than Vega, they have nearly the same apparent brightness in the sky. You can conclude that Rigel must be emitting a lot more light than Vega.


The changing shape of the Big Dipper
100,000 years ago the Big Dipper had a different shape.
Arrows show amounts and directions of the Big Dipper stars' proper motions across the sky.
100,000 years in the future, the Big Dipper's shape will have changed further.

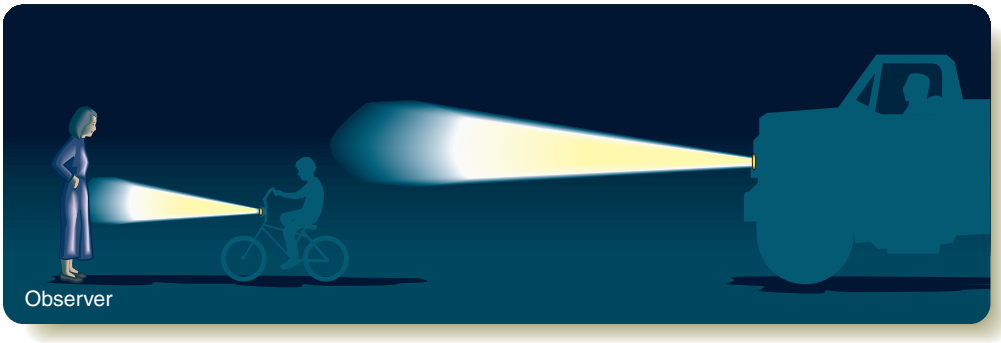
▲ **Figure 9-3** Proper motion refers to slow changes seen from Earth in the positions of individual stars and the shapes of constellations, caused by the stars' real movements through space.

This comparison of Vega with Rigel shows how astronomers can use proper motion to identify stars that are likely to be nearby. You have seen this effect if you watch birds in flight. Birds generally move at roughly the same true speed, but distant birds seem to move slowly across the sky whereas a nearby bird flits quickly across your field of view. If a star has a small proper motion, it is probably a distant star, but a star with a large proper motion is probably quite close.

9-2 Apparent Brightness, Intrinsic Brightness, and Luminosity

If you see a light on a dark highway, it is hard to tell how powerful it really is. It could be the brilliant headlight on a distant truck or the dim headlight on a nearby bicycle (**Figure 9-4**). Obviously, how bright an object appears depends not only on how much light it emits but also on its distance.

A sixth-magnitude star just visible to your eye looks faint, but its apparent magnitude doesn't tell you how luminous it really is. (Look back to Sections 2-1d and 2-1e to review the magnitude system used to describe the apparent brightness of stars.) Now that



◀ **Figure 9-4** To judge the true brightness of a light source, you need to know how far away it is. The headlight on a distant truck might appear as bright as the light on a nearby bicycle, giving you no clue about their real distances.

you know how to find the distance to stars, you can use those distances to determine the **intrinsic brightness** of the stars. Intrinsic means "belonging to the thing," so the intrinsic brightness of a star refers to the total amount of light the star emits.

9-2a Brightness and Distance

When you look at a light, your eyes respond to the visual-wavelength photons falling on your eye's retina. The apparent brightness you perceive is related to the flux of energy entering your eye. **Flux** is measured in watts per square meter (W/m^2); in other words, the energy in watts (joules per second) falling on an area of one square meter.

The flux (apparent brightness) of a light source is determined by the inverse square law. You first encountered the inverse square relation in Chapter 5 (Figure 5-5), where you learned that Newton guessed that the strength of gravitational force decreases inversely with the square of distance because he had already determined that light behaves this way. In other words, Newton found that if he doubled his distance from a light source, its brightness (the flux he received) fell by a factor of 2^2, which equals four times. If he tripled his distance, the brightness (flux) fell by a factor of 3^2, or nine times. The flux of light reaching you is inversely proportional to the square of your distance from the source.

You can see that if you know both the apparent magnitude of a star (expressing the star's flux received on Earth) and its distance, you can use the inverse square law to correct for distance and learn the intrinsic brightness of the star. Astronomers express intrinsic brightness using special magnitude units described in the next section.

9-2b Absolute Visual Magnitude and Distance

If all stars were the same distance from Earth, you could compare one with another and easily tell which was emitting more light and which less. Of course, the stars are scattered at different distances, and you can't move them into a lineup for comparison. If, however, you know the distance to a star, you can use the inverse square relation to calculate the brightness the star would have at some standard distance. Astronomers have adopted 10 pc as the standard distance and refer to the **absolute visual magnitude (M_V)** of a star as the apparent visual magnitude it would have if it were 10 pc away. The absolute visual magnitude is therefore an expression of the intrinsic brightness of the star.

To find a star's absolute visual magnitude, you begin by measuring its apparent visual magnitude (in other words, its brightness), a relatively easy task. Then, you need the distance to the star. If the star is nearby, you can measure its parallax and calculate the distance. Once you know the distance, you can use the magnitude–distance formula that you will learn in the next section to correct the apparent visual magnitude for the distance and find the absolute visual magnitude.

The subscript V in the symbol for absolute visual magnitude reminds you that it is a visual magnitude that only includes the wavelengths of light a human eye can see. Other magnitude systems are based on other parts of the electromagnetic spectrum such as infrared and ultraviolet.

How does the Sun stack up against other stars? The Sun is tremendously bright in the sky, but it is also very nearby. Its absolute visual magnitude is just +4.8. In other words, if the Sun were only 10 pc from Earth (about 33 ly, which is not a great distance in astronomy), it would look no brighter than the faintest star in the handle of the Little Dipper.

The intrinsically brightest stars known have absolute visual magnitudes of about −8.3, which means that such a star 10 pc from Earth would be more than 25 times as bright as Venus at its brightest. Such stars have intrinsic brightness 13 magnitudes brighter than the Sun, which means they are emitting more than 100,000 times more light at visible wavelengths than the Sun (look back to Table 2-1). In contrast, the intrinsically faintest stars have absolute visual magnitudes of about +16, which is 11 magnitudes fainter than the Sun, meaning they are emitting 25,000 times less visible light than the Sun. You now have a sense of proportion about the wide range of star characteristics.

9-2c Calculating Absolute Magnitude

If you know a star's apparent visual magnitude and its distance, you can calculate its absolute visual magnitude. The **magnitude–distance formula** that allows this calculation relates apparent visual magnitude, m_V, distance in parsecs, d, and absolute visual magnitude, M_V:

KEY EQUATION 9-2a Absolute Magnitude and Distance

$$m_v - M_v = -5 + 5 \log(d)$$

The expression *log* means logarithm to the base 10. Sometimes it is convenient to rearrange the equation and write it in this form:

KEY EQUATION 9-2b Absolute Magnitude and Distance

$$d = 10^{(m_v - M_v + 5)/5}$$

The two equations are really just two versions of one equation, so you can use whichever form is most convenient in a given problem. If you know the distance, the first form of the equation is convenient, but if you are trying to find the distance, the second form of the equation is better.

Example: Famous Star Polaris is 133 pc from Earth and has an apparent visual magnitude of +2.0. What is its absolute visual magnitude? ***Solution:*** A pocket calculator tells you that log(133) equals 2.12, so, solving for M_V tells you that the absolute visual magnitude of Polaris is −3.6. If it were only 10 pc from Earth, it would dominate the night sky.

9-3d Luminosity (Total Energy Output)

The **luminosity (*L*)** of a star is the total energy the star radiates in 1 second. Hot stars emit a great deal of ultraviolet light that you can't see, and cool stars emit mostly infrared light. Absolute visual magnitude includes only visible radiation, so astronomers need to make a correction—sometimes quite large—accounting for the invisible energy. Then they can calculate the total luminosity of the star from its absolute magnitude.

Astronomers often express luminosities in solar units, writing 2.5 $L_\odot$ to represent a luminosity 2.5 times greater than the Sun's. To find the luminosity of a star in watts, you can just multiply by the luminosity of the Sun in those units, 3.8×10^{26} W. For example, Famous Star Aldebaran has a luminosity of about 170 times the Sun's, which corresponds to about 6.5×10^{28} W.

The most luminous stars emit at least a billion times more energy per second than the least luminous. Clearly, the family of stars contains some interesting characters.

Practicing SCIENCE

How can two stars look the same in the sky but have dramatically different luminosities? This is the kind of question a scientist might ask herself as an exercise to think through what is already known about the relationships among the apparent brightnesses, luminosities, and distances of stars.

The farther away a star is, the fainter it looks—that is just the inverse square law. Famous Stars Vega and Rigel have about the same apparent visual magnitude, which means your eyes receive the same amount of light from them. But parallax observations from the *Hipparcos* satellite reveal that Rigel is 34 times farther away than Vega, so Rigel must be more than 1000 times as luminous as Vega.

Distance is often the key to understanding the brightness of stars, but temperature can also be important. Ask another question to review how we know what we know about stars: ***Why must astronomers make a correction in converting the absolute visual magnitude of very hot or very cool stars into luminosities?***

9-3 Stellar Spectra

Observations of spectral lines give you information about what types of atoms are in the atmospheres of the Sun, planets, and stars, as you learned in Chapter 7. In the late 1800s and early 1900s, when astronomers were making the first careful studies of stellar spectra, the obvious differences observed between spectra of different stars were thought to indicate that stars have a wide range of compositions.

In Section 8-1e, you read about the story of Cecilia Payne, who made use of information from atomic physics and the new field of quantum mechanics to reinterpret the spectra of the Sun and stars. Payne's calculations showed that more than 90 percent of the atoms in the Sun and other stars must be hydrogen and most of the rest are helium (look back to Table 8-1). Payne discovered that (1) the chemical compositions of the Sun and other stars are very similar, and (2) spectra actually provide information mostly about the temperatures of stars, meaning that stars with similar spectra must have similar temperatures.

9-3a The Balmer Thermometer

Look back to the Chapter 7 Concept Art: Atomic Spectra and review how spectral absorption lines are produced in stellar atmospheres. The light that forms a spectrum comes from the transparent gases of the photosphere and atmosphere, so the spectrum is direct information only about those outer layers.

One of the main methods Payne first developed for using spectra to determine star temperatures is now called the *Balmer thermometer.* Normal star surface temperatures range from about 2300 K to about 50,000 K; for comparison, the surface temperature of the Sun is 5780 K. From information about blackbody radiation and Wien's law in Section 7-2c, you already know how astronomers estimate a star's temperature by using its color, but the spectral lines of hydrogen at wavelengths visible to the human eye—called the *Balmer lines*—combined with a few other spectral lines, can give much greater precision.

The Balmer thermometer works because the strength of the Balmer lines depends on the temperature of the star's surface layers. Both hot and cool stars have weak Balmer lines, but medium-temperature stars have strong Balmer lines. That is because Balmer absorption lines are produced only by atoms with electrons in the second energy level.

If a star is cool, there are few violent collisions between atoms to excite the electrons, so the electrons of most atoms are in the ground state, not the second level. Electrons in the ground state can't absorb photons in the Balmer series. As a result, you should expect to find only weak Balmer absorption lines in the spectra of cool stars. On the other hand, in the surface layers of hot stars there are many violent atomic collisions that can excite electrons to high energy levels or even ionize some atoms by knocking electrons completely out of them. In that situation also, there are few hydrogen atoms with their electrons in the

second orbit to form Balmer absorption lines. Therefore, hot stars, like cool stars, have weak Balmer absorption lines.

In stars of an intermediate temperature—roughly 10,000 K—collisions have just the right strength and rate to excite large numbers of electrons into the second energy level. Hydrogen gas at that temperature absorbs photons very well at the wavelengths of the Balmer lines and produces strong (dark) spectral lines.

Theoretical calculations can predict how strong the Balmer lines should be for stars of various temperatures. Such calculations are the key to finding temperatures from stellar spectra. The curve in **Figure 9-5a** shows the strength of the Balmer lines for various stellar temperatures. You can see from the graph that a star with Balmer lines of a certain strength might have either of two temperatures, one high and one low. How do you know which is right? You have to examine spectral lines of substances other than hydrogen to be sure you have the correct temperature.

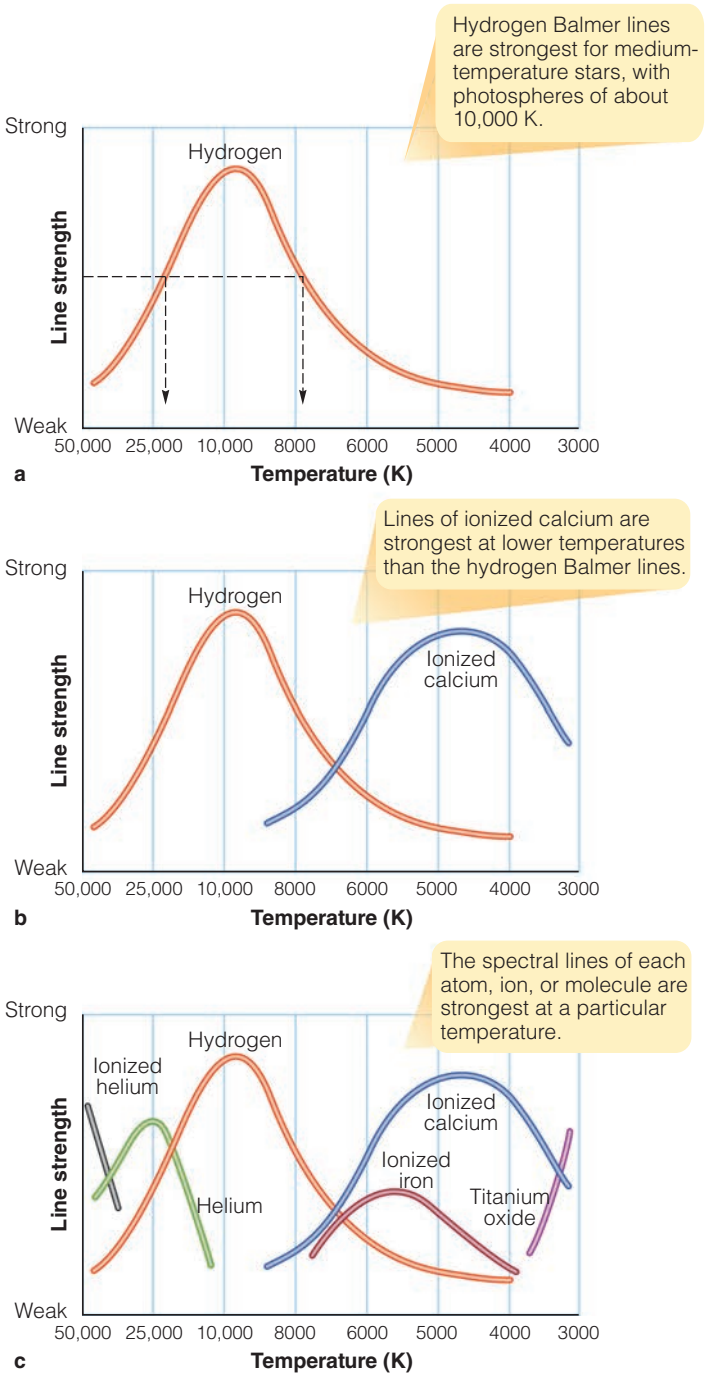
Spectral line strengths for elements other than hydrogen also depend on temperature, but the temperature at which the lines reach their maximum strength differs for each element (**Figure 9-5b**). If you add a number of chemical elements to your graph, you will have a powerful aid for finding the temperatures of stars (**Figure 9-5c**). For example, if you find medium-strength Balmer lines and strong helium lines in a star's spectrum, you can conclude that it has a temperature of about 20,000 K. If instead the star has weak hydrogen lines, medium-strong lines of ionized calcium, and strong lines of ionized iron, you would assign it a temperature of about 5500 K to 6000 K, similar to that of the Sun.

The spectra of stars cooler than about 3500 K contain dark bands produced by molecules such as titanium oxide (TiO). Because of their structure, molecules can absorb photons at many wavelengths, producing numerous closely spaced spectral lines that blend together to form bands. These molecular bands appear in the spectra of only the coolest stars because, as mentioned before, molecules in cool stars are not subject to violent collisions that would break them apart.

9-3b Temperature Spectral Types

During the 1890s, astronomers at Harvard Observatory invented the first widely used system for classifying stellar spectra. One of those scientists, Annie J. Cannon, personally inspected the spectra of more than 250,000 stars. Spectra were first classified into groups labeled A through Q, but some of those groups were later dropped, merged with others, or reordered. The final classification scheme includes seven major temperature **spectral types**, or classes, still used today: O, B, A, F, G, K, M. (Generations of astronomy students have remembered the spectral type sequence using the mnemonic, "Oh, Be A Fine Girl [Guy], Kiss Me." You can probably invent a better one yourself.)

This stellar **spectral sequence** is a temperature sequence. O stars are the hottest, and temperature decreases along the sequence to M stars, the coolest. For finer definition, astronomers



▲**Figure 9-5** The strength of spectral lines can tell you the temperature of a star, a principle discovered by Cecila Payne in the 1920s. (a) Hydrogen Balmer lines alone are not enough because they give two possible answers: Strong Balmer lines mean the star's temperature is around 10,000 K, but weak or moderate-strength Balmer lines can be produced by a very hot or a very cool star. (b) Adding another atom to the diagram helps, and (c) adding many atoms and molecules to the diagram creates a precise aid to determine the temperatures of stars.

divide each spectral type into ten subtypes. For example, spectral type A consists of the subtypes A0, A1, A2, …, A8, A9. Next come F0, F1, F2, and so on. The Sun is not just a G star, but a G2 star. This finer division gives a star's temperature to a precision of about 3 percent. **Table 9-1** breaks down some of the

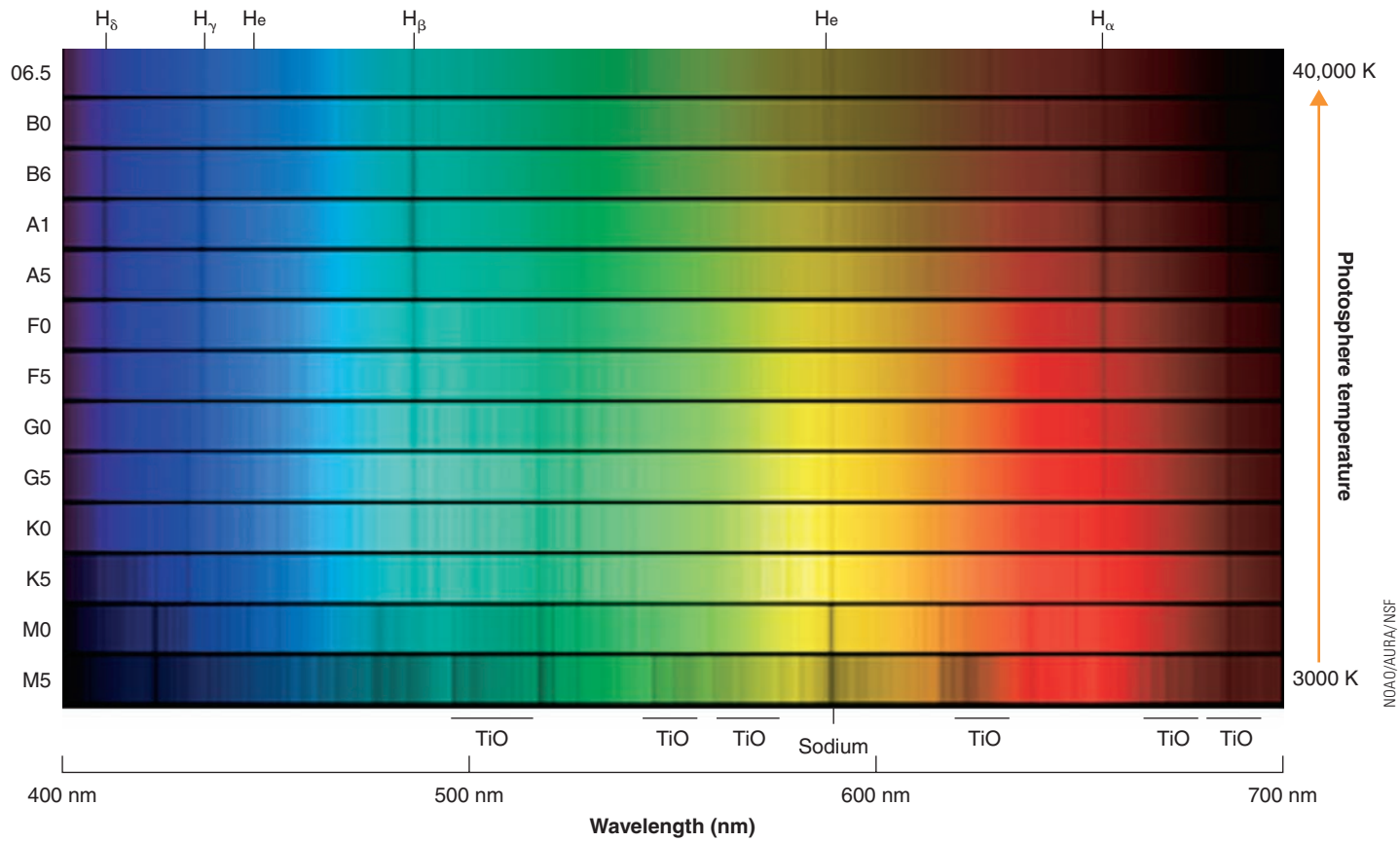
Table 9-1 Temperature Spectral Types

Spectral Class	Approximate Temperature (K)	Hydrogen Balmer Lines	Other Spectral Features	Naked-Eye Example
O	40,000	Weak	Ionized helium	λ Orionis (O8)
B	15,000	Medium	Neutral helium	Achernar (B3)
A	8500	Strong	Ionized calcium weak	Sirius (A1)
F	6600	Medium	Ionized calcium weak	Canopus (F0)
G	5500	Weak	Ionized calcium medium	Sun (G2)
K	4100	Very weak	Ionized calcium strong	Aldebaran (K5)
M	3000	Very weak	TiO strong	Betelgeuse (M2)

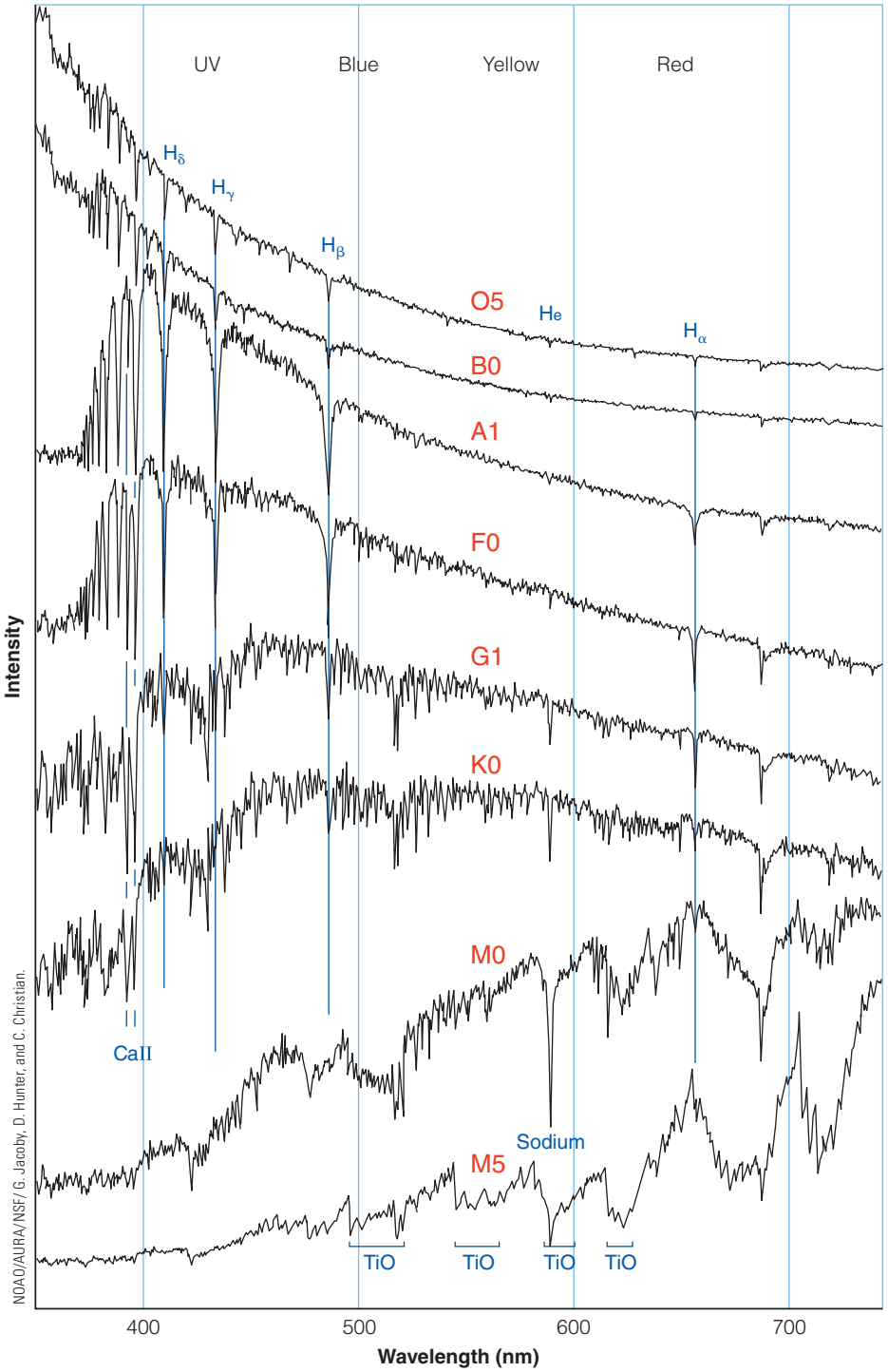
information contained in Figure 9-5c and presents it according to spectral type. For example, if a star has weak Balmer lines and lines of ionized helium, it must be an O star.

Thirteen stellar spectra are arranged in **Figure 9-6** from the hottest at the top to the coolest at the bottom. You can easily see in those spectra how the strength of spectral lines depends on temperature; the Balmer thermometer you have just learned about is especially obvious. The hydrogen Balmer lines are strongest in A stars that have middle-range temperatures, weak in hotter stars (O and B), and weak in cooler stars (F through M).

Although these color photos of spectra are attractive, astronomers today normally do not work with spectra in the form of images. Rather, as you learned in Section 7-3, spectra are usually displayed as graphs of intensity versus wavelength that show dark absorption lines as dips in the graph (**Figure 9-7**). Such graphs allow more detailed analysis than photographs.



▲**Figure 9-6** A compilation of color photographs of stellar spectra ranging from hot O stars at the top to cool M stars at the bottom: The hydrogen Balmer lines are strongest at spectral type A0, but the two closely spaced lines of sodium in the yellow are strongest for very cool stars. Helium lines appear only in the spectra of the hottest stars. Notice that the helium line visible in the top spectrum has nearly, but not exactly, the same wavelength as the sodium lines visible in cooler stars. Bands produced by the molecule titanium oxide (TiO) are strong in spectra of the coolest stars.

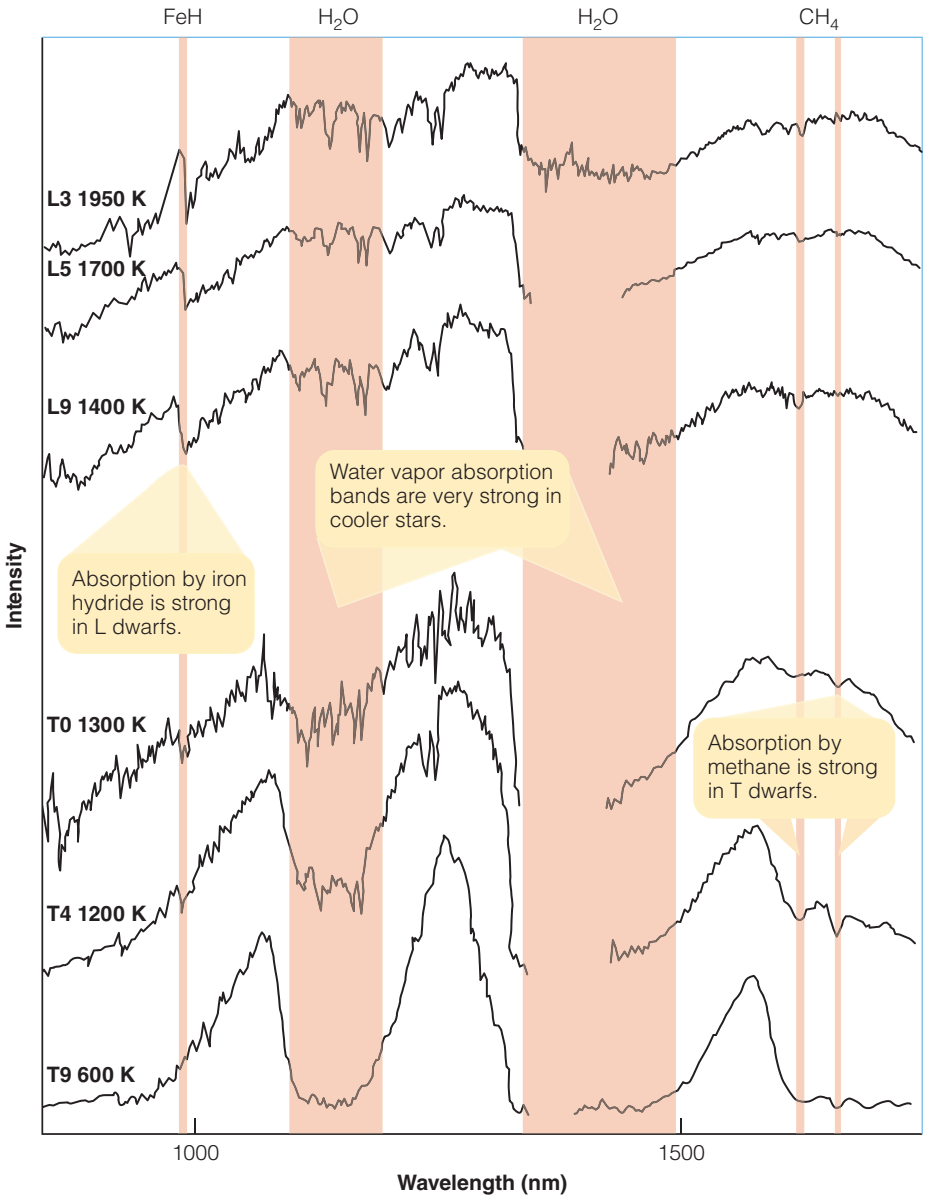


◀ **Figure 9-7** Modern digital spectra are usually represented as graphs of intensity versus wavelength, with dark absorption lines appearing as sharp dips in the curves. The hottest stars are at the top and the coolest at the bottom. Hydrogen Balmer lines are strongest at spectral type A0, whereas lines of ionized calcium (Ca II) are strongest in K stars. Titanium oxide (TiO) bands are strongest in M stars. Compare these spectra with Figures 9-5c and 9-6.

Notice, for example, that the overall graphs have the shapes of blackbody curves with spectral lines superimposed. The wavelength of maximum intensity is in the infrared for the coolest stars and in the ultraviolet for the hottest stars. Look carefully at these graphs, and you can see that helium lines are visible only in the spectra of the hottest classes and titanium oxide bands only in the coolest. Two lines of ionized calcium that you can see at wavelengths around 390 nm increase in strength from A to K and then decrease from K through M. Because the strengths of these spectral lines depend on temperature, it requires only a few moments to study a star's spectrum and estimate its temperature.

Now you can learn something new about some of the Famous Stars. Sirius is an A1 star, and Vega is an A0 star. Thus, they have nearly the same temperature and color, and both have strong hydrogen Balmer lines in their spectra. The bright orange star in Orion is Betelgeuse, a cool M2 star, whereas blue-white Rigel in Orion is a hot B8 star. Polaris, the North Star, is an F8 star a bit hotter than our Sun, and Alpha Centauri, the closest visible star to the Sun, is a yellow-white G2 star just like the Sun.

The study of spectral types is more than a century old, but astronomers continue to discover and define new types. The **L dwarfs**, found in 1998, are cooler and fainter than M stars. They are understood to be objects smaller than stars but larger than planets and have been named **brown dwarfs**. You will learn more about them in a later chapter. The spectra of M stars contain bands produced by metal oxides such as titanium oxide, but L dwarf spectra contain bands produced by more fragile molecules such as iron hydride (FeH). The **T dwarfs**, discovered in 2000, are an even cooler and fainter type of brown dwarf than L dwarfs. Their spectra show absorption by methane (CH_4) and water vapor (**Figure 9-8**). In 2011, astronomers using highly sensitive infrared detectors and large ground-based telescopes discovered a class of objects with temperatures below 500 K that are labeled **Y dwarfs**.



◀ **Figure 9-8** These six infrared spectra show the differences between L dwarfs and T dwarfs. Spectra of M stars (Figure 9-7, *bottom*) show titanium oxide bands (TiO), but L and T dwarfs are so cool that other molecules, such as iron hydride (FeH), water (H_2O), and methane (CH_4), dominate their spectra.

9-4 Star Sizes

Now that you know the luminosities of stars, you are ready to find their sizes, which are expressed usually as radii or diameters. Recall that astronomers can't see stars as disks; almost all stars look simply like points of light, no matter how big the telescope. Using interferometry techniques (look back to Section 6-5d), the sizes of a few stars have been measured, and surface features have been distinguished on a very small number of stars including Famous Star Betelgeuse. Nevertheless, there is a straightforward way to find the sizes of stars. If you know a star's temperature and luminosity, you can calculate its radius.

9-4a Luminosity, Radius, and Temperature

To find the size of a star from its luminosity, you first need to recall from Section 7-2c the two factors that determine the energy output of blackbodies such as stars: surface area and temperature. You can eat dinner by candlelight because a candle flame has a small surface area. Although the flame is very hot, it cannot radiate much heat; it has a low luminosity. However, if the candle flame were 12 feet tall, it would have a very large surface area from which to radiate, and although it might be no hotter than a normal candle flame, its luminosity would drive you from the table (**Figure 9-9**).

In a similar way, a hot star may not be very luminous if it has a small surface area, but it can be highly luminous if it has a large surface area from which to radiate. On the other hand, even a cool star can be luminous if it has a large surface area.

Now you can practice the simple calculation of the luminosity of a star with known temperature and size. Recall from the discussion of the Stefan-Boltzmann law of blackbody radiation (Eq. 7-1) that the amount of energy emitted per second from each square meter of the surface of an opaque object such as a star equals σT^4. Thus, the star's luminosity can be written as its surface area in square meters times the amount it radiates from each square meter:

$$L = (\text{surface area}) \times \sigma T^4$$

Because a star is a sphere, you can use this geometric formula: surface area $= 4\pi R^2$. Then, the luminosity is:

$$L = 4\pi R^2 \sigma T^4$$

If you express luminosity, radius, and temperature in proportion to the Sun, there is a simpler form of the equation:

KEY EQUATION 9-3 Luminosity, Radius, and Temperature

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \left(\frac{T}{T_\odot}\right)^4$$

Note that astronomers use the symbol ⊙ to refer to the Sun. Thus, $L_\odot$ refers to the luminosity of the Sun, $T_\odot$ refers to the temperature of the Sun, and so on.

Krafft/Science Source

◀**Figure 9-9** Molten lava pouring from a volcano is not as hot as a candle flame, but a lava flow has more surface area and radiates more energy than a candle flame. Approaching a lava flow without protective gear is dangerous.

Example 1: Suppose you want to find the luminosity of a star that has ten times the Sun's radius but a surface temperature only half that of the Sun. How luminous is it? ***Solution:*** Substituting those values into the equation,

$$\frac{L}{L_\odot} = \left(\frac{10}{1}\right)^2\left(\frac{1}{2}\right)^4 = \left(\frac{100}{1}\right)\left(\frac{1}{16}\right) = 6.25$$

This star therefore has 6.25 times the Sun's luminosity.

Now, you can use the same formula to determine the size of a star relative to the Sun if you know its luminosity and temperature in solar units. ***Example 2:*** Suppose you measure the brightness and parallax of a star and determine its intrinsic brightness. Its spectrum shows it has twice the Sun's temperature, which allows you to correct the intrinsic brightness to include nonvisible radiation (look back to Section 9-3d) and calculate that its total luminosity is 40 $L_\odot$. What is the radius of the star relative to the Sun's radius? ***Solution:*** If you put the values for luminosity and temperature relative to the Sun into the formula, you can find the radius:

$$\frac{40}{1} = \left(\frac{R}{R_\odot}\right)^2\left(\frac{2}{1}\right)^4$$

Solving for the radius, you get:

$$\left(\frac{R}{R_\odot}\right)^2 = \frac{40}{2^4} = \frac{40}{16} = 2.5$$

So the radius is:

$$\frac{R}{R_\odot} = \sqrt{2.5} \approx 1.6$$

Therefore, a star with 40 times the Sun's luminosity and 2 times the Sun's temperature must be about 1.6 times larger than the Sun.

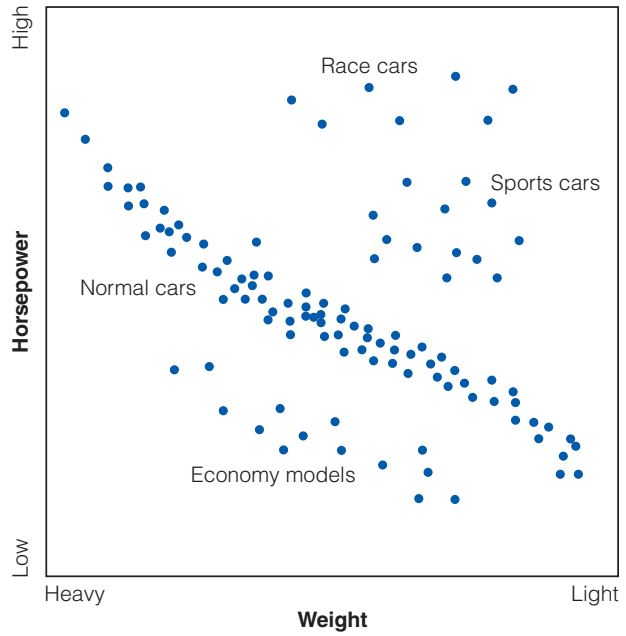
9-4b The H–R Diagram

The **Hertzsprung–Russell (H–R) diagram**, which is named after its originators—Netherlands astronomer Ejnar Hertzsprung and U.S. astronomer Henry Norris Russell—is a graph that separates the effects of temperature and surface area on stellar luminosities and enables astronomers to sort and classify stars according to their sizes. The H–R diagram is one of the most important information displays used in astronomy. In later chapters this diagram will help you learn about the life cycles of stars.

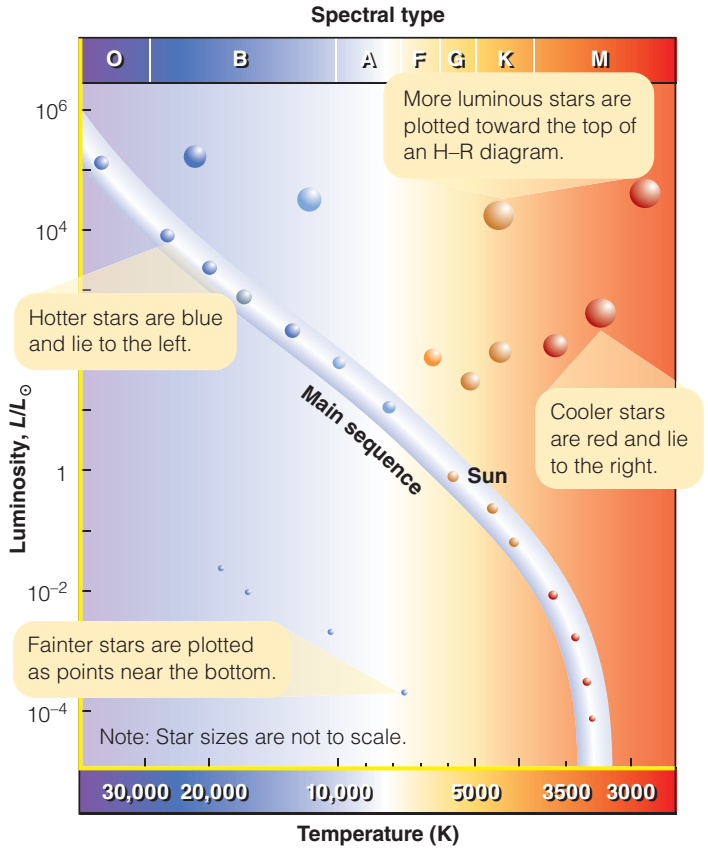
Before you explore the details of the H–R diagram, try looking at a similar diagram you might use to describe and classify automobiles. You can plot a diagram such as **Figure 9-10**, showing horsepower versus weight for various makes of cars. In general, the more a car weighs, the more horsepower it has. Most cars fall somewhere along the sequence of ordinary cars, running from heavy, high-powered cars at the upper left to light, low-powered models at the lower right. You might call this the main sequence of cars. But some cars have much more horsepower than normal for their weight—the sports and racing models—and lie higher in the diagram. Other cars, the economy models, have less power than normal for cars of the same weight and fall lower in the diagram. Just as this diagram can be used to sort cars into engine-power groups, so the H–R diagram sorts stars into groups according to size.

The H–R diagram is a graph with luminosity on the vertical axis and temperature on the horizontal axis. A star is represented on the graph by a point that marks its luminosity and its temperature. The H–R diagram in **Figure 9-11** also contains a scale of spectral type across the top. As you now know, a star's temperature determines its spectral type, so you can use either spectral type or temperature as the horizontal axis.

In an H–R diagram, the location of a point tells you a great deal about the star it represents. Points near the top of the



▲ **Figure 9-10** You could analyze automobiles by plotting their horsepower versus their weight and thus reveal relationships among various models. The properties of most cars would lie somewhere along the main sequence of "normal" cars.



▲ **Figure 9-11** In an H–R diagram, a star is represented by a dot at a position that shows the star's luminosity and temperature. The background color in this diagram indicates the temperature of the stars. The Sun is a yellow-white G2 star. Most stars including the Sun have properties along the main-sequence strip running from hot high-luminosity stars at upper left to cool low-luminosity stars at lower right.

diagram represent very luminous stars, and points near the bottom represent stars with very low luminosities. Points near the right edge of the diagram represent very cool stars, and points near the left edge of the diagram represent very hot stars. The artist who drew the H–R diagram in Figure 9-11 used color to represent stellar temperatures. As you learned in Section 7-2b regarding Wien's law, red stars are cool, and blue stars are hot. (Astronomers use H–R diagrams so often that they usually skip the words "the point that represents the star." Rather, they will say that a star is located in a certain place in the diagram, although the location of a star in the H–R diagram has nothing to do with the location of the star in space. Furthermore, a star is said to "move" in the H–R diagram as it ages and its luminosity and temperature change, but "motion" in this diagram has nothing to do with the star's motion in space.)

9-4c Giants, Supergiants, and Dwarfs

The **main sequence** is represented in Figure 9-11 as the curved line with dots for stars plotted along it, running from upper left to lower right of the H–R diagram. It includes roughly 80 percent of all stars. As you might expect, hot main-sequence stars are more luminous than cool main-sequence stars.

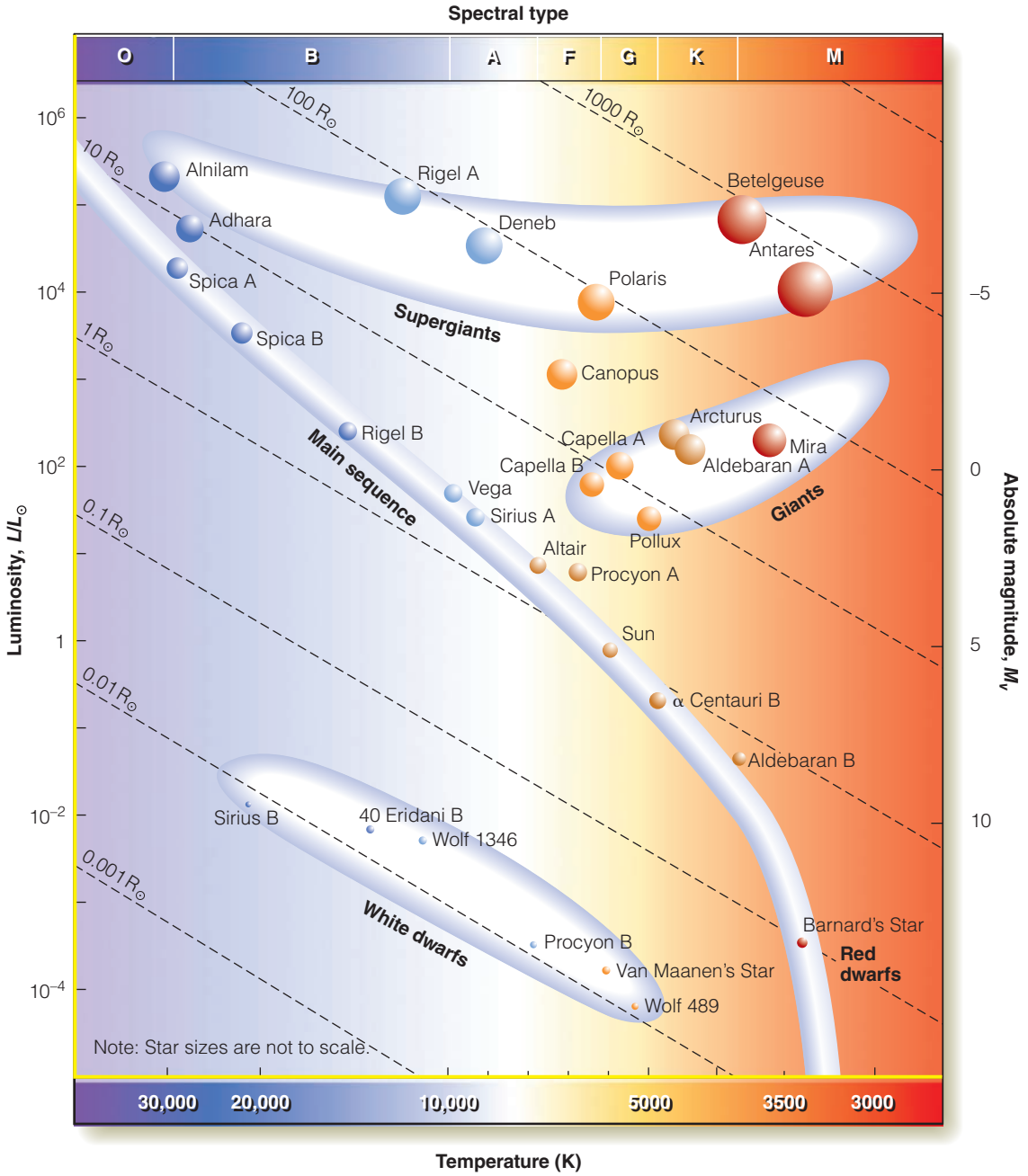
Notice in the H–R diagram that some cool stars lie above the main sequence. Although they are cool, these **giants** are luminous, and that means they must have more surface area and therefore larger diameters than main-sequence stars of the same temperature. To give you a sense of proportion, these giant stars are roughly 10 to 100 times larger than the Sun. Some of the **supergiants** at the very top of the H–R diagram have more than a thousand times the Sun's diameter. To visualize the size of the largest stars, imagine that the Sun is the size of a tennis ball. Then, the largest supergiants would be the size of a sports stadium.

At the bottom of the H–R diagram are the "economy models," stars that are very low in luminosity because they are very small. **Red dwarfs** at the lower end of the main sequence, in the lower right corner of the H–R diagram, are not only relatively small, they are also cool, and that gives them low luminosities. In contrast, the **white dwarfs** lie in the lower left of the H–R diagram and are lower in luminosity than you would expect, given their high temperatures. That must mean they are very small, smaller than red dwarfs. Again, if the Sun were the size of a tennis ball, white dwarfs would be the size of grains of sand. Although some white dwarfs are among the hottest stars known, they are so small they have very little surface area from which to radiate, and that makes their luminosities low.

The equation you have just studied that relates luminosity, temperature, and radius of a star can be used to draw precise lines of constant radius across the H–R diagram. These lines slope down and to the right across the diagram because cooler stars are

fainter than hotter stars of the same size. **Figure 9-12** plots the luminosities and temperatures of the Famous Stars and some other well-known stars along with lines of constant radius. For example, locate the line labeled 1 $R_\odot$ (1 solar radius) and notice that it passes through the point representing the Sun. Any star whose point is located along this line has a size equal to the Sun's. Next, look at the rest of the stars along the main sequence. They range from a tenth the size of the Sun to about ten times as large. Even though the main sequence slopes dramatically down to the right across the diagram, most main-sequence stars are similar in size. In contrast, the white dwarfs at the lower left of the diagram are extremely small—only 1/100 the diameter of the Sun, about the size of Earth—and, as you have already learned, the giants and supergiants at the upper right are extremely large compared to main-sequence stars.

Notice the great range of sizes among stars. Supergiants are 100,000 times larger than the tiny white dwarfs. Now you can understand why Famous Star Rigel is so much more luminous than



▲**Figure 9-12** An H–R diagram showing the luminosity and temperature of many well-known stars. (Individual stars that orbit each other are designated A and B, for example Spica A and Spica B.) The dashed lines are lines of constant radius; star sizes on this diagram are not to scale. If the Sun were the size of a tennis ball, the largest supergiants would be the size of a sports stadium and white dwarfs the size of grains of sand.

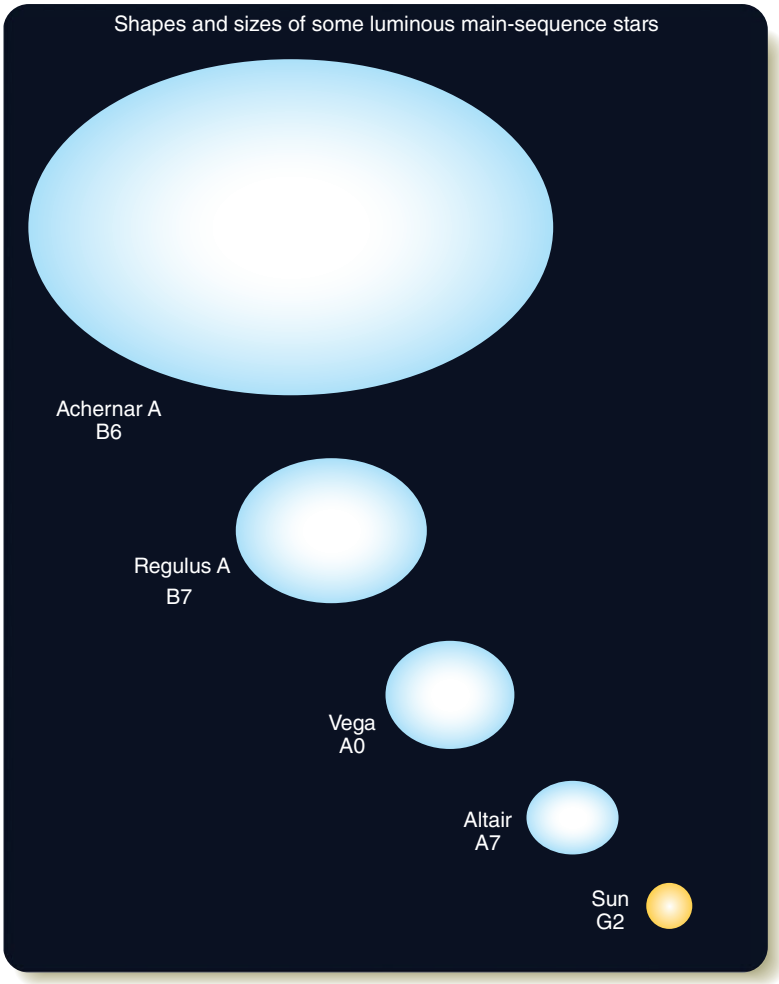
Vega. They have nearly the same temperature, but Rigel is a supergiant, with a much larger surface area than Vega's from which to radiate. Famous Star Betelgeuse in Orion is also a supergiant. If Betelgeuse replaced the Sun at the center of our Solar System, it would swallow up Venus, Earth, Mars, and Jupiter. The largest stars known have radii of about 7 AU; if one of them replaced the Sun, it would extend nearly to the orbit of Saturn.

9-4d Interferometer Observations of Star Diameters

Is there any way to check the diameters predicted by the H–R diagram? One way is to use interferometers, for example the Center for High Angular Resolution Astronomy (CHARA) Array on Mount Wilson near Los Angeles. There, six 1-m telescopes observing at optical and near-infrared wavelengths combine their light to produce the resolving power of a virtual telescope 330 m (1080 ft) in diameter. Such interferometers can resolve the diameters of large, bright stars.

You can see Famous Star Vega high in the sky on summer and fall evenings. It is an A0 star, and observations with the CHARA array confirmed previous interferometry measurements that Vega is about 2.6 times larger in diameter than the Sun. The observations also reveal that Vega is spinning about twice a day compared to once a month for the Sun, giving the star a flattened shape with an equatorial diameter 20 percent larger than its polar diameter (Figure 9-13).

Other interferometer observations show that, as expected, upper-main-sequence stars are larger than the Sun. Altair (Alpha Aquilae), an A7 star, has an equatorial diameter about twice the Sun's, and Regulus A (Alpha Leonis A), a B7 star, is about four times larger than the Sun. Achernar A (Alpha Eridani A), a B3 star, is even larger. All three of these stars are flattened by their rapid rotation, as you can see in Figure 9-13.



▲**Figure 9-13** Observations with interferometers can resolve the size and shape of some nearby stars. The stars of the upper main sequence are indeed larger than the Sun, as predicted by the H–R diagram. The examples shown here are flattened by rapid rotation, but most stars rotate slower and are more nearly spherical. On the scale of this diagram, the supergiant Betelgeuse would have a diameter of about 7 meters (23 feet). (An A added to a star's name indicates that it is the most luminous component of a binary star system.)

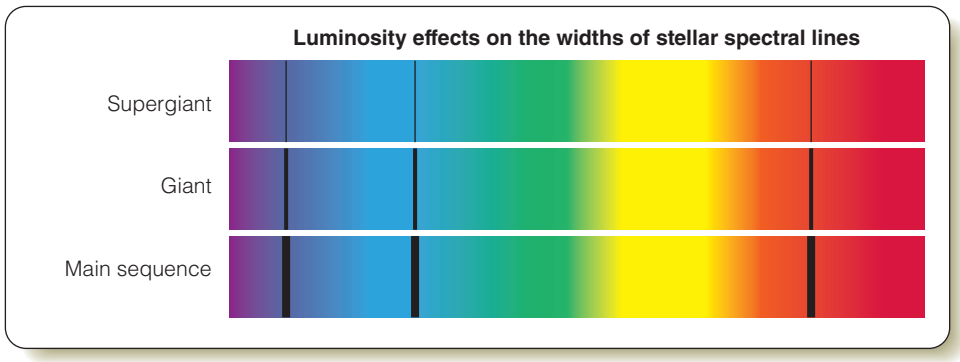
Interferometer observations of giant and supergiant stars such as Betelgeuse (M2) show that they are also as large as predicted. Betelgeuse has a diameter of about 10 AU, 1000 times the size of the Sun. Thus, direct measurements with interferometers confirm the sizes of stars estimated from their temperatures and luminosities using the Stefan-Boltzmann law. Stars really do range from roughly one-hundredth to more than a thousand times the size of the Sun.

9-4e Luminosity Spectral Classification

Spectra of stars don't contain just clues about their temperature and composition; you can also use a star's spectrum to determine whether it is a main-sequence star, a giant, or a supergiant, even if it is too far away for interferometry. The larger a star is, the less dense its atmosphere is, and that affects the widths of spectral lines.

Atoms collide often in a dense gas, their energy levels become distorted, and their spectral lines are broadened. Hydrogen Balmer absorption lines provide a good example (Figure 9-14). Balmer lines in the spectrum of a main-sequence star are broad because the star's atmosphere is relatively dense and the hydrogen atoms collide often. In the spectrum of a giant star, the spectral lines are narrower because a giant star's atmosphere has lower density than a main-sequence star of the same temperature, and the hydrogen atoms collide less often. The Balmer lines in the spectrum of a supergiant star are even narrower.

Thus, you can look at a star's spectrum and tell roughly how big it is. Size categories derived from spectra are called **luminosity classes** because the size of the star is the dominating factor in determining luminosity. The luminosity classes are represented by Roman numerals: Ia for luminous supergiants, Ib for supergiants, II for luminous giants, III for giants, IV for low-luminosity



◀ **Figure 9-14** These model spectra show how the widths of spectral lines reveal a star's luminosity class. Supergiants have very narrow spectral lines, and main-sequence stars have broad lines. Certain spectral lines are more sensitive to this effect than others; careful inspection of a star's spectrum can determine its luminosity classification.

giants (called *subgiants*), V for main-sequence stars (sometimes referred to in this context as *dwarfs*), and VI for low-luminosity main-sequence stars (called *subdwarfs*).

This classification scheme allows luminous supergiants (Ia) such as Rigel (Beta Orionis) to be distinguished from "ordinary" supergiants (Ib) such as Polaris (Alpha Ursae Minoris). The star Adhara (Epsilon Canis Majoris) is a luminous giant (II), Aldebaran (Alpha Tauri) is a giant (III), and Altair (Alpha Aquilae) is a subgiant (IV). Sirius (Alpha Canis Majoris) and Vega (Alpha Lyrae) are main-sequence stars (V), and Mu Cassiopeiae is a subdwarf (VI). When you describe a star, its luminosity class appears after the spectral type; for example, the Sun is G2 V, a G2 main-sequence star. White dwarfs don't enter into this classification because, as you will learn later, they are actually remnants of stars, and their spectra are very different from any of the other types. Notice that several of the Famous Stars (Rigel, Polaris, Aldebaran) are supergiants or giants; next time you look at the North Star, remind yourself that it is a supergiant.

Luminosity classification is subtle and not very accurate, but it is important in modern astronomy. As you will see in the next section, luminosity classes provide a way to estimate the distance to stars that are too far away to have measurable parallaxes.

9-4f Spectroscopic Parallax

Astronomers can measure the stellar parallax of nearby stars, but most stars are too distant to have measurable parallaxes. Their distances can be estimated from the star's spectral type, luminosity class, and apparent magnitude with a procedure called **spectroscopic parallax**. Spectroscopic parallax is a potentially confusing term because it does not involve measurement of parallax shifts, but it is a method to find the distance to the star.

Spectroscopic parallax relies on the location of the star in the H–R diagram. If you record the spectrum of a star, you can determine its spectral type, and that tells you its horizontal location in the H–R diagram. You can also determine its luminosity class by looking at the widths of its spectral lines, and that allows you to estimate the star's vertical location in the diagram. Once you plot the point that represents the star in the H–R diagram, you can read its absolute magnitude based on stars with the same luminosity and temperature classes that are close enough to have measurable parallaxes. The last step is finding the star's distance by comparing its apparent and absolute magnitudes.

For example, Famous Star Betelgeuse has an apparent magnitude of about +0.4. It is classified M2 Ib (M2 temperature class, supergiant luminosity class). You can find Betelgeuse plotted in Figure 9-12 at an absolute magnitude of about −6.0. The difference between Betelgeuse's apparent and absolute magnitudes, $m_V - M_V$, is therefore 0.4 minus (−6.0), or 6.4, so its distance calculated from the magnitude–distance formula (Eq. 9-2b) is about 190 pc. A combination of parallax measurements made by the *Hipparcos* satellite and radio telescopes yields a distance of 197 pc with an uncertainty of 24 percent, so the result derived from the spectroscopic parallax method is pretty good. Obviously a real measurement of the parallax is better, but for distant stars spectroscopic parallax is often the only way to find their distance.

Practicing SCIENCE

What is the evidence that there are giant stars much bigger than the Sun? A scientist knows that such a basic, broad statement about stars should not be accepted uncritically, but must be firmly supported by observations.

Stars exist that have the same spectral type as the Sun but are clearly more luminous because we know their distances and therefore their luminosities. For example, component Ab of the binary star Capella (Alpha Aurigae) is a G1 star with an absolute visual magnitude, based on its distance, of +0.2. Because it is a G1 star, it has almost exactly the same temperature as the Sun, but its absolute magnitude is 4.6 magnitudes brighter than the Sun's. A magnitude difference of 4.6 corresponds to a ratio of about 70, so Capella must be about 70 times more luminous than the Sun. If it has the same surface temperature as the Sun but is 70 times more luminous, then it must have a surface area 70 times greater than the Sun's. Because the surface area of a sphere is proportional to the square of the radius, Capella Ab must have a radius a bit more than 8 times larger than the Sun's. That is clear observational evidence that Capella is a giant star.

In Figure 9-12, you can see that Procyon B is a white dwarf slightly warmer than the Sun but about 10,000 times less luminous. Try another exercise at forming conclusions based on simple observations and principles of physics: ***Why do astronomers conclude that white dwarfs must be much smaller than the Sun?***

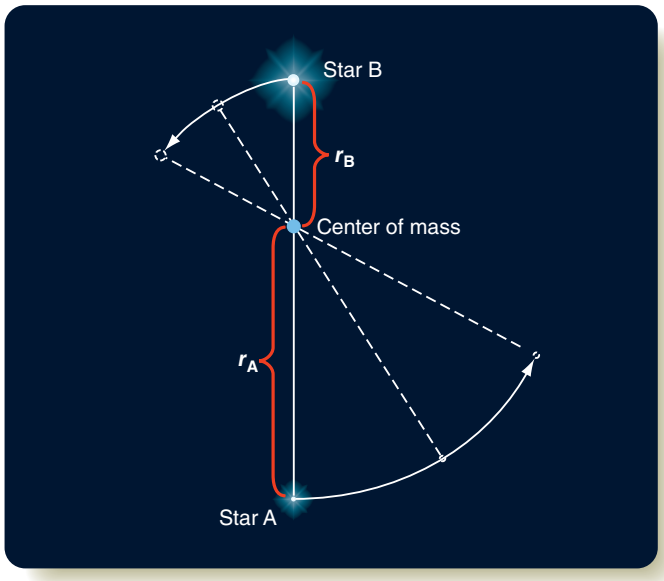
9-5 Star Masses—Binary Stars

To understand how stars form, generate energy, and evolve, you first need to find their masses; that is, how much matter they contain. As you will learn in the next several chapters, a star's mass determines its life history. And the key to measuring star masses is gravity. Matter produces a gravitational field, and you can determine how much matter a star contains if you watch another object being affected by that star's gravitational field. In other words, the most straightforward way to find the masses of stars is to study **binary stars**, pairs of stars that orbit each other.

9-5a Binary Stars in General

The Chapter 5 Concept Art: Orbits illustrates orbits by having you imagine a cannonball fired from a high mountain. If Earth's gravity didn't act on the cannonball, it would follow a straight-line path and leave Earth forever; instead, Earth's gravity forces the cannonball to follow a curved path around Earth—an orbit. When two stars orbit each other, their mutual gravitation pulls them away from straight-line paths and causes them both to follow closed orbits around a point between the stars (Figure 9-15).

To find the total mass of a binary star system, you need to know the sizes of the orbits and the orbital period, which is the length of time the stars take to complete one orbit. The smaller the orbits are and the shorter the orbital period is, the stronger the stars' gravity must be to hold each other in orbit and, therefore, the more massive the stars must be.



▲ **Figure 9-15** As stars in a binary star system revolve around each other, the line connecting them always passes through the center of mass, and the more massive star is always closer to the center of mass.

9-5b Calculating the Masses of Binary Stars

Kepler's third law of orbital motion works only for the planets in our Solar System. When Newton realized that mass produces the gravitational attraction that causes orbital motion, he made that law into a general principle. Newton's version of Kepler's third law (Eq. 5-2) applies to any pair of objects that orbit each other. The total mass of the two objects is related to the average distance a between them and their orbital period P. In other words, the period and total size of the two stars' orbits tells you the sum of their masses. If the masses of the two stars are symbolized by M_A and M_B, then:

KEY EQUATION 9-4 Kepler's Third Law for Stellar Masses

$$M_A + M_B = \frac{a^3}{P^2}$$

In that formula, the average separation between the stars a is expressed in AU, orbital period P in years, and the star masses M_A and M_B in solar masses. If you can measure the average distance in AU between the two stars and their orbital period in years, the sum of the masses of the two stars equals a^3/P^2.

Example 1: If you observe a binary star system with a period of 32 years and an average separation of 16 AU, what is the total mass? ***Solution:*** The total mass is $16^3/32^2$, which equals 4 solar masses.

Consider what happens if you apply the binary star formula to the motion of a planet in our Solar System. In other words, pretend that the Sun and your chosen planet form a binary star system. The total mass is effectively just the mass of the Sun, 1 $M_\odot$, because all of the planets, even Jupiter, have tiny masses relative to the Sun. Then, the binary star formula becomes $P^2 = a^3$, which is Kepler's third law (Section 4-3e; Eq. 4-1).

What about the masses of the individual stars? Each star in a binary system moves in its own orbit around the system's center of mass, the balance point of the system. If one star is more massive than its companion, then the more massive star is closer to the center of mass and travels in a smaller orbit, while the lower-mass star whips around in a larger orbit. The ratio of the masses of the stars M_A/M_B equals r_B/r_A, the inverse of the ratio of the radii of the orbits. If one star has an orbit twice as large as the other star's orbit, then it must be half as massive. The total size and period of the stars' orbits give you their total mass, and the relative sizes of the two orbits give you the ratio of their masses. Combined, that is enough information to let you determine the individual masses of each star.

Example 2: Suppose, in the previous example, the distance of Star A from the center of mass is 12 AU, and Star B's is 4 AU. What are their individual masses? ***Solution:*** The ratio of the

Chains of Inference

How do scientists determine something they can't measure directly?

Sometimes scientists cannot directly observe the things they really want to study, so they must construct chains of inference that connect observable parameters to the unobservable quantities they want to know. You can't measure the mass of a star directly, so you must find a way to use what you can observe, orbital period and angular separation, to figure out step-by-step the quantities you need to calculate the mass.

Consider another example. Geologists can't measure the temperature and density of Earth's interior directly. There is no way to drill a hole to Earth's center and lower a thermometer or recover a sample. However, the speed of vibrations from a distant earthquake depends on the temperature and density of the rock they pass through. Geologists can't measure the speed of the vibrations deep inside Earth, but they can measure the delays in the arrival times at different locations on the surface, and that allows them to work their way back to the speed and, finally, the temperature and density.

Chains of inference can be nonmathematical. Biologists studying the migration of whales can't always follow individual whales for years at a time, but they can observe them feeding and mating in different locations; take into consideration food sources, ocean currents, and water temperatures; and construct a chain of inference that leads back to the seasonal migration pattern for the whales.

This chapter contains a number of chains of inference. Almost all fields of science use chains of inference. When you can link the observable quantities step-by-step to the final conclusions, you gain a strong insight into the nature of science.

USGS

The San Andreas fault: A chain of inference connects our understanding of earthquakes to our understanding of conditions deep inside Earth.

masses must be 12:4, which equals 3:1. Now, what two numbers add up to 4 and have the ratio 3:1? Star B must have 3 solar masses, and Star A must have 1 solar mass.

Figuring out the masses of stars of a binary star system normally has some complications. The orbits of the two stars may be eccentric, and although the orbits lie in the same plane, that plane can be tipped at an unknown angle to your line of sight, distorting the observed shapes of the orbits. Astronomers must find ways to correct for these distortions. Also, converting the apparent angular sizes of the two orbits into true sizes requires knowing the distance to the system. Finding the masses of binary stars requires a number of steps to get from what you can observe to what you really want to know, the masses. Constructing such sequences of steps is an important part of science (**How Do We Know? 9-1**).

Although there are many different kinds of binary stars, three types are especially useful for determining stellar masses. These are discussed separately in the next section.

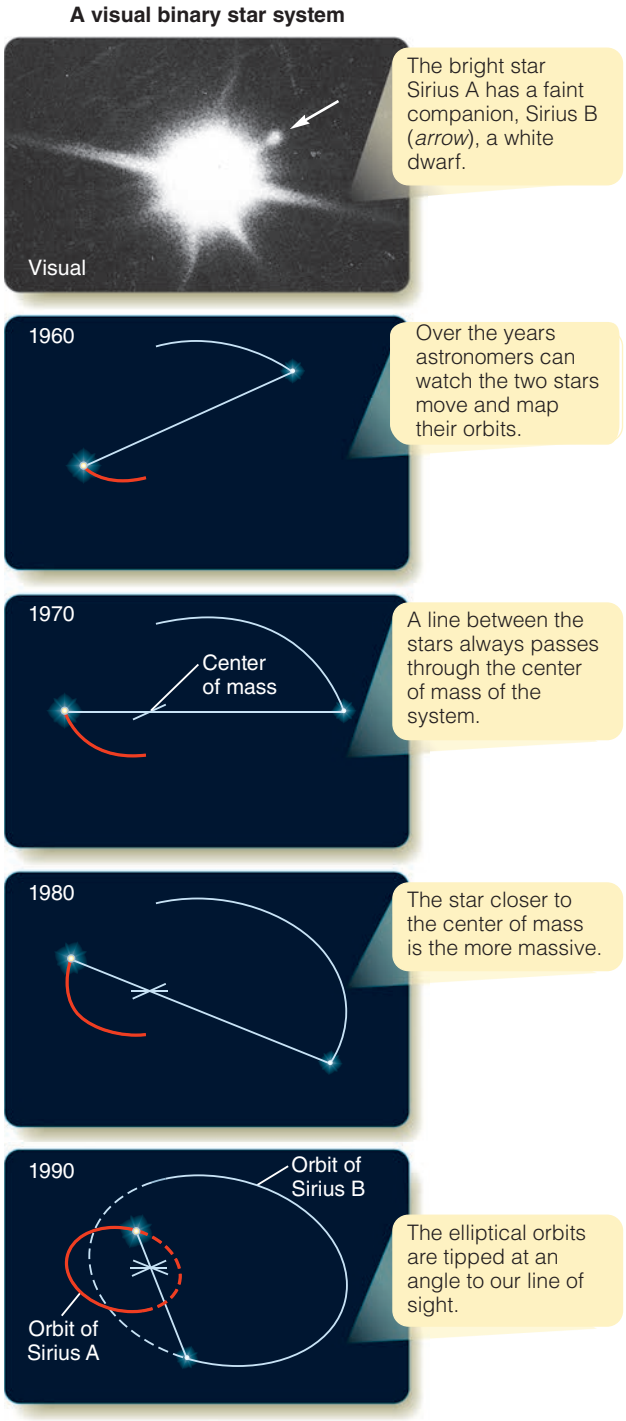
9-5c Three Types of Binary Systems

Although there are many different kinds of binary stars, three types are especially important for determining masses and other properties of stars. Studying binary stars is also preparation for the much more difficult problem of finding planets around stars other than our Sun, because a star with a planet orbiting around it is like a binary star system with one very small component. Each type of binary star system corresponds to a different technique for detecting planets, as you will learn in a later chapter.

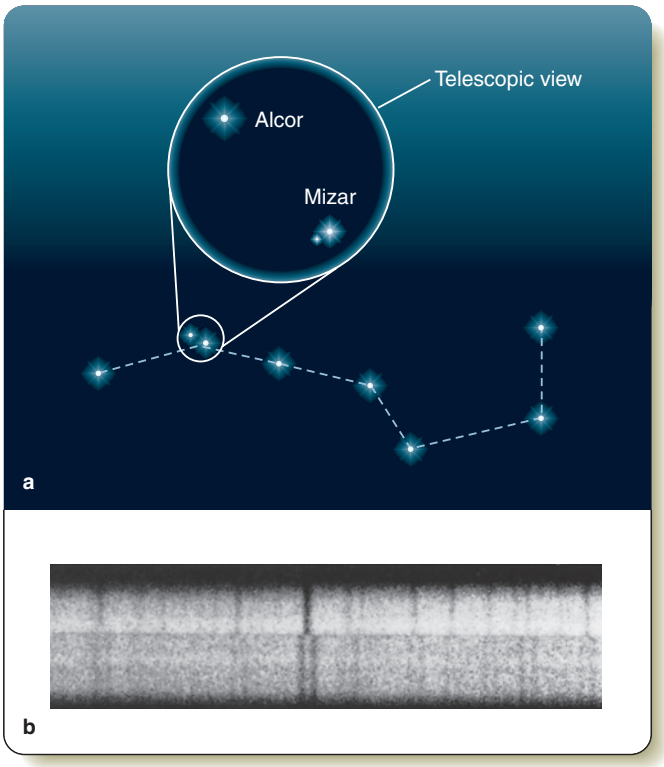
In a **visual binary system**, the two stars are separately visible in the telescope and astronomers can watch the stars orbit each other over periods of years or decades, as the series of illustrations of the binary components of Famous Star Sirius in **Figure 9-16** demonstrates. From that, astronomers can find the orbital period and, if the distance of the system from Earth can be found, the size of the orbits. That is enough to find the masses of the stars. (Note, however, that many visual binaries have such large orbits that their orbital periods are hundreds or thousands of years, and in those cases astronomers have not yet seen an entire orbit.)

Some binary stars orbit so close to each other they are not visible as separate stars. If the stars in a binary system are close together, a telescopic view, limited by diffraction and atmospheric seeing, shows a single point of light. Such systems can't be analyzed as a visual binary. Only by looking at a spectrum, which is formed by light from both stars and contains spectral lines from both, can astronomers tell that there are two stars present and not one. Such a system is called a **spectroscopic binary system**. Familiar examples of spectroscopic binary systems are the stars Mizar and Alcor in the handle of the Big Dipper (**Figure 9-17**).

Figure 9-18 shows a pair of stars orbiting each other; the circular orbit appears elliptical because you see it nearly edge-on. If this were a true spectroscopic binary system, you would not see


A visual binary star system
The bright star Sirius A has a faint companion, Sirius B (arrow), a white dwarf.
Visual
1960
Over the years astronomers can watch the two stars move and map their orbits.
1970
Center of mass
A line between the stars always passes through the center of mass of the system.
1980
The star closer to the center of mass is the more massive.
1990
Orbit of Sirius B
Orbit of Sirius A
The elliptical orbits are tipped at an angle to our line of sight.

▲**Figure 9-16** Measuring the orbital motion of Sirius A and Sirius B allows calculation of their individual masses.



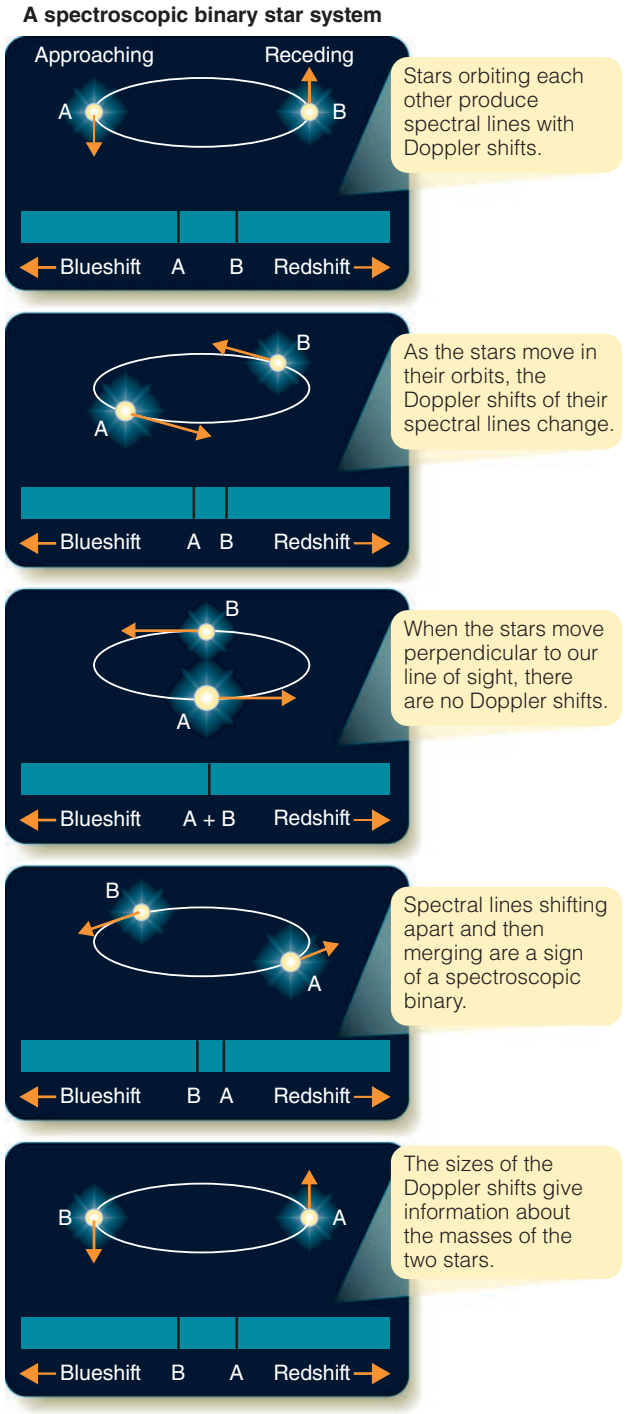
The Observatories of the Carnegie Institution of Washington

▲**Figure 9-17** (a) At the bend of the handle of the Big Dipper lies a pair of stars, Mizar (Zeta Ursae Majoris) and Alcor (80 Ursae Majoris). Through a telescope you can discover that Mizar has a fainter companion and so is a member of a visual binary system. Adaptive optics observations reveal a faint close companion of Alcor, not pictured in this diagram. (b) Spectra of Mizar sometimes show single spectral lines (*top strip*) and other times show double lines (*bottom strip*), indicating that the main star of the visual binary system is itself a spectroscopic binary system rather than a single star.

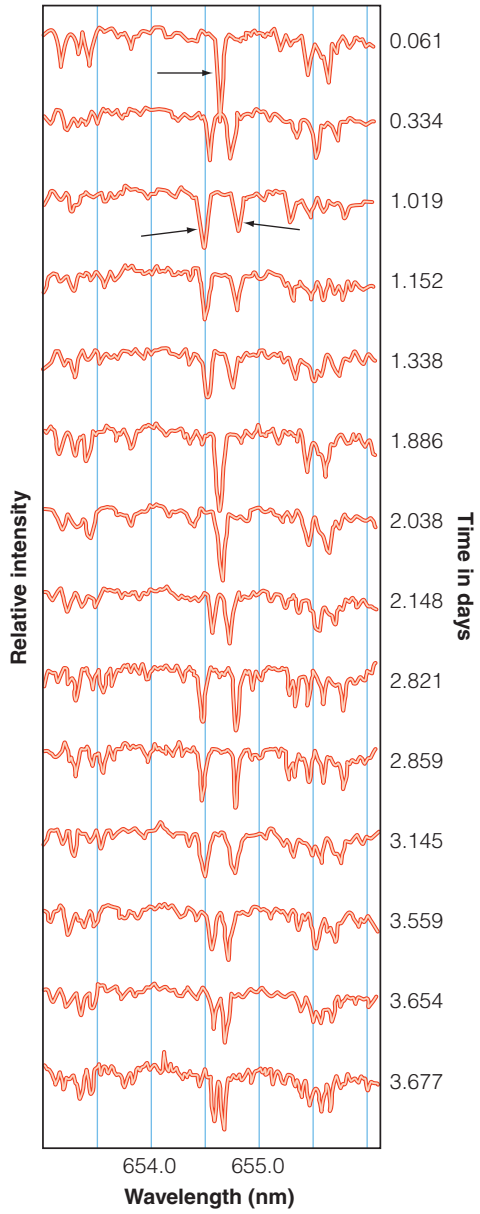
the separate stars, but, as the stars move in their orbits, they alternately approach toward and recede from Earth and their spectral lines are Doppler shifted alternately toward blue and then red wavelengths (Section 7-3b). Noticing pairs of spectral lines moving back and forth across each other in a series of spectra would alert you that you are observing a spectroscopic binary (**Figure 9-19**).

Although spectroscopic binaries are very common, they are not as useful as visual binaries for determining star masses. The orbital period can be measured easily, but because the stars are not resolved, the true size of the orbits can't be determined because there is no way to know the angle at which the orbits are tipped to our view from Earth. That means the true masses of a spectroscopic binary cannot be calculated; the best that astronomers can do in this situation is determine lower limits to the masses.

For some spectroscopic binary systems, however, the orbit of the stars is nearly edge-on to Earth, so from our view the stars cross in front of each other each orbit. Although the system looks like a single point of light, when one star moves in front of the other star, the star "at the back" is partly or totally eclipsed and some of the light is blocked. As a result the total brightness of the system decreases temporarily. Such a binary is called an **eclipsing binary system**. Note that eclipsing binaries are not really different beasts from spectroscopic binaries. If a spectroscopic binary happens to be oriented edge-on to our view from Earth, we see eclipses and call it an eclipsing binary.

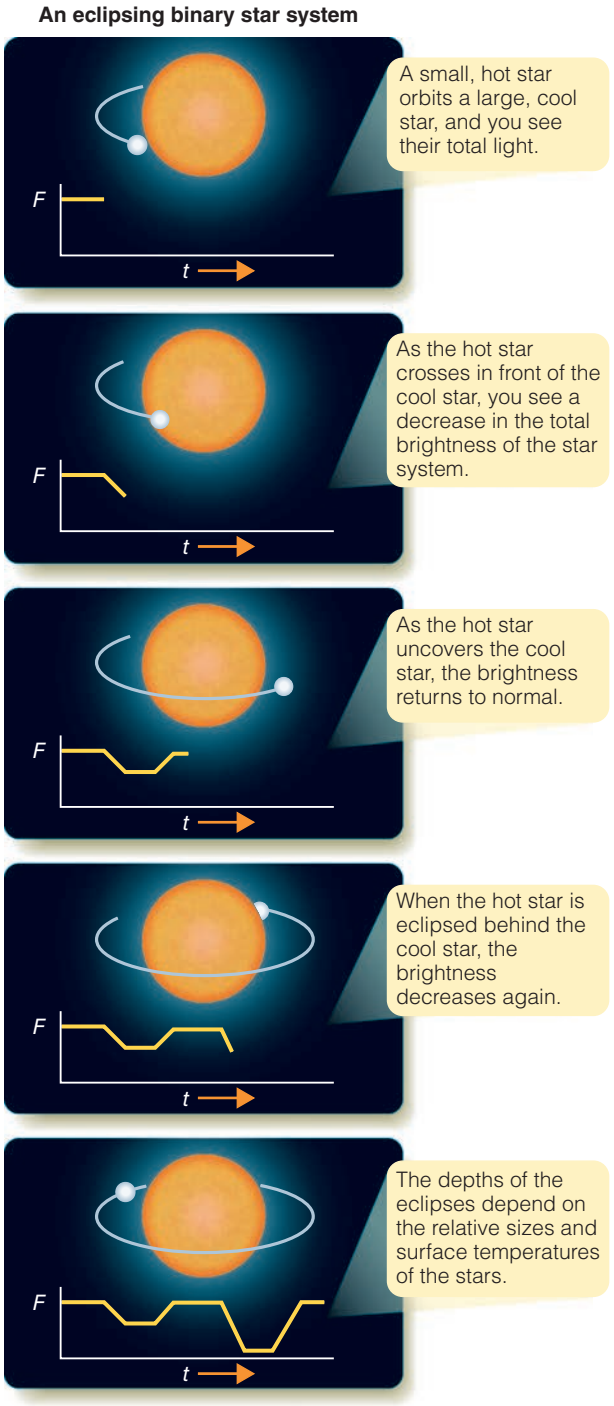


▲**Figure 9-18** From Earth, a spectroscopic binary looks like a single point of light, but the changing Doppler shifts in its spectrum reveal the orbital motion of the two stars (see also Figure 9-17b).



▲**Figure 9-19** Fourteen spectra of the star HD 80715 are shown here as graphs of intensity versus wavelength. A single spectral line (arrow in top spectrum) splits into a pair of spectral lines (indicated by arrows in third spectrum from the top), which then merge and split apart again. These changing Doppler shifts reveal that HD 80715 is a spectroscopic binary.

Figure 9-20 shows a smaller star moving in an orbit around a larger star, first eclipsing the larger star and then being eclipsed as it moves behind. The resulting variation in the brightness of the system is shown as a graph of brightness versus time, which is called a **light curve**. Remember, you can't see the individual stars in an eclipsing system. Cover the stars in the bottom panel of the figure with your fingers and look only at the complete light curve. If you saw such a light curve, you would immediately recognize that what looks like a single star in the sky is actually an eclipsing binary.


An eclipsing binary star system
A small, hot star orbits a large, cool star, and you see their total light.
As the hot star crosses in front of the cool star, you see a decrease in the total brightness of the star system.
As the hot star uncovers the cool star, the brightness returns to normal.
When the hot star is eclipsed behind the cool star, the brightness decreases again.
The depths of the eclipses depend on the relative sizes and surface temperatures of the stars.
F
t

▲ **Figure 9-20** From Earth, an eclipsing binary looks like a single point of light, but changes in brightness reveal that two stars are eclipsing each other. A series of measurements of the system's brightness (called the *light curve*) shown here as flux versus time, combined with measurements of Doppler shifts in the spectra, can reveal the sizes and masses of the individual stars.

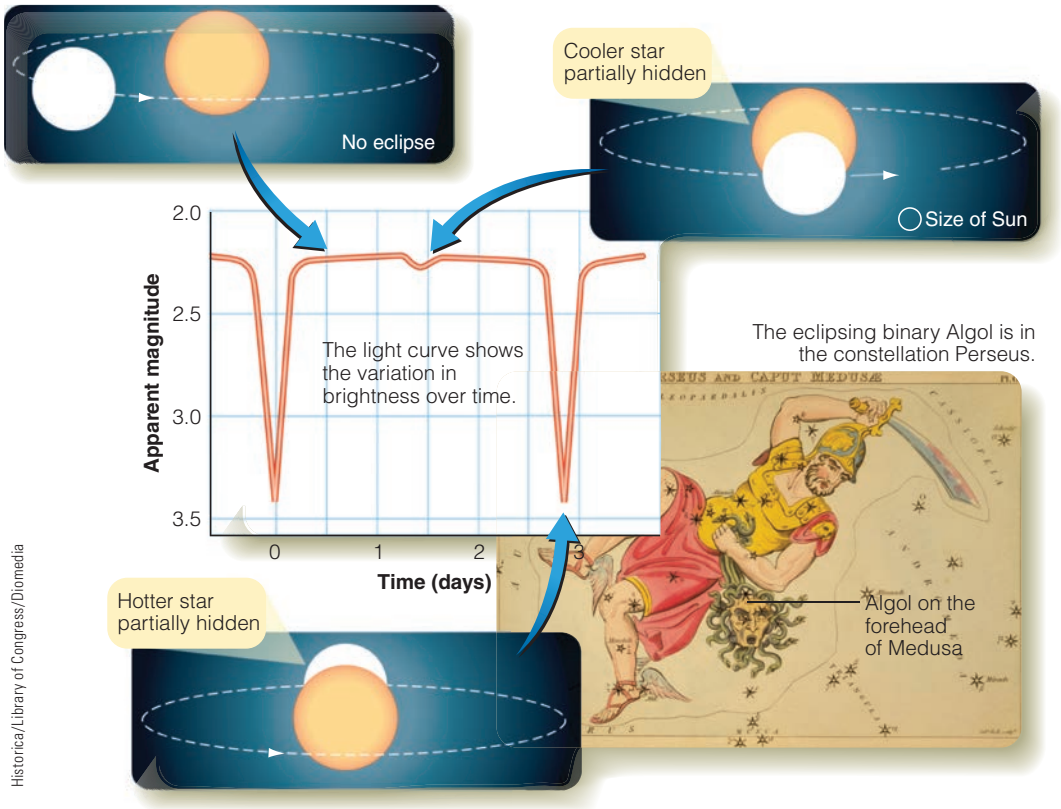
Algol (Beta Persei) is one of the best-known eclipsing binaries because its eclipses are visible to the naked eye. Normally, its visual magnitude is about +2.2, but its brightness drops to +3.4 (a factor of 3 dimmer) during eclipses, which occur every 68.8 hours. Although the nature of the star as an eclipsing binary was not recognized until 1783, its periodic dimming has probably been known since ancient times. Algol comes from the Arabic for "the demon," and it is associated in constellation mythology with the severed head of Medusa, the sight of whose serpentine locks turned mortals to stone (**Figure 9-21**). Indeed, in some accounts, the variable star Algol is the winking eye of the demon.

The light curves of eclipsing binary systems can be difficult to analyze, but they contain lots of information. **Figure 9-22a** shows the light curve of a binary system in which the stars have dark spots on their surfaces, analogous to enormous sunspots, and are so close to each other that their shapes are distorted. The light curve of this system allows crude maps of the two stars to be derived (**Figure 9-22b**).

Once the light curve of an eclipsing binary system has been accurately observed, astronomers can construct a chain of inference that leads to the masses of the two stars. They can find the orbital period easily from the light curve. Then, they can get spectra showing the Doppler shifts of the two stars, which directly yield their orbital speeds because no correction is needed for the inclination of the orbits; the orbits must be nearly edge-on, or there would not be eclipses. From the orbital speeds and period, astronomers can find the sizes of the orbits, at which point they have all the information needed to determine the masses of the stars.

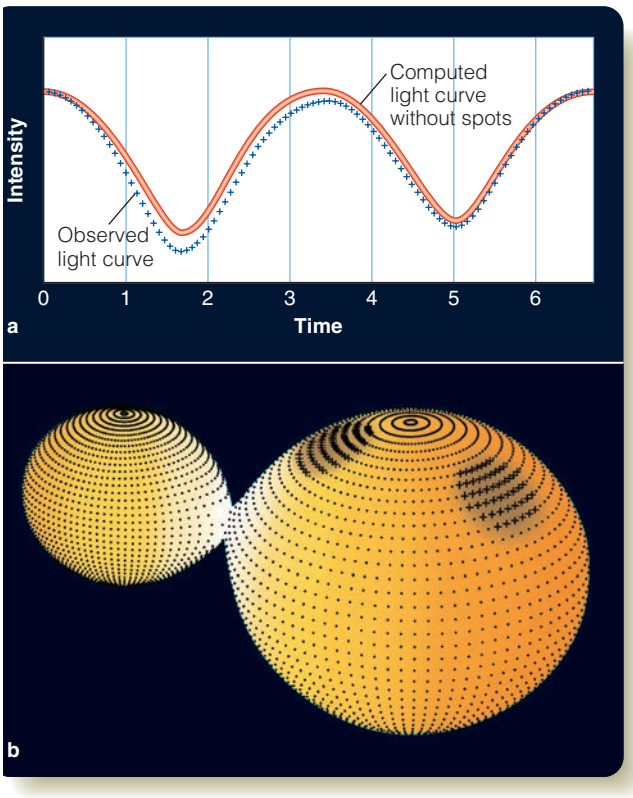
Previously in this chapter you learned that luminosity and temperature can be used to determine the radii of stars. Eclipsing binary systems provide a way to check those calculations by allowing direct measurement of the sizes of a few stars. The light curve shows how long it takes for the stars to cross in front of, or be hidden by, each other. Multiplying these time intervals by the orbital speeds measured from the Doppler shifts gives the diameters of the stars. For example, if it takes 3000 seconds for the small star to disappear behind the big star while moving 50 km/s, the small star must be 150,000 km in diameter. As you learned in the previous section, there are complications resulting from the inclinations and eccentricities of orbits, but often these effects can be taken into account. Thus, observations of an eclipsing binary system can directly tell you not only the masses of the two stars but also their diameters.

From the study of binary stars, astronomers have found that the masses of stars range from 0.08 solar mass to more than 100 solar masses. The most massive stars ever found in a binary system have about 150 solar masses. A few other stars may be even more massive, but they are not members of binary systems, so astronomers must estimate their masses from models based on their luminosities, temperatures, and compositions.



Historica/Library of Congress/Diomedia

◀ **Figure 9-21** The eclipsing binary Algol consists of a hot B star and a cooler K star. The eclipses are partial, meaning that neither star is completely hidden during eclipses. The orbit here is drawn as if the cooler star were stationary. The relative size of the Sun is indicated by the small open circle.



▲ **Figure 9-22** The observed light curve of the binary star VW Cephei (pronounced *SEF-ee-eye; lower curve*) shows that the two stars are so close together their gravity distorts their shapes. Slight variations in the light curve reveal the presence of transient dark spots at specific places on the star's surface. The upper curve shows what the light curve would look like if there were no spots.

Practicing **SCIENCE**

When astronomers look at the light curve for an eclipsing binary system that has total eclipses, how do they know which star is hotter? Start by assuming that the two stars in an eclipsing binary system are not the same size, so they can be labeled the larger star and the smaller star. When the smaller star moves behind the larger star, you lose the light coming from the total area of the small star.

When the smaller star moves in front of the larger star, it blocks off light from the same amount of area on the larger star. In both cases, the same area (the same number of square meters) is hidden from your sight. That means the amount of light lost during an eclipse depends only on the temperature of the hidden surface because temperature is what determines how much a single square meter can radiate per second. When the surface of the hotter star is hidden, the brightness will fall dramatically, but when the surface of the cooler star is hidden, the brightness will not fall as much. So, you can look at the light curve and point to the deeper of the two eclipses and say, "That is where the hotter star is behind the cooler star."

Now consider a related problem, determining the diameters of the stars in an eclipsing binary system. ***How would you use the light curve of a system that has total eclipses to find the ratio of the two stars' diameters?***

9-6 A Census of the Stars

You have learned how to find the luminosities, temperatures, sizes, and masses of stars, and now you can put all those data together (**How Do We Know? 9-2**) to paint a family portrait of the stars. As in any family portrait, both similarities and differences are important clues to the history of the family. As you begin to understand, while studying the next few chapters, how stars are born, evolve, and die, keep in mind a simple question: What is the average star like? Answering that question is both challenging and illuminating.

9-6a Surveying the Stars

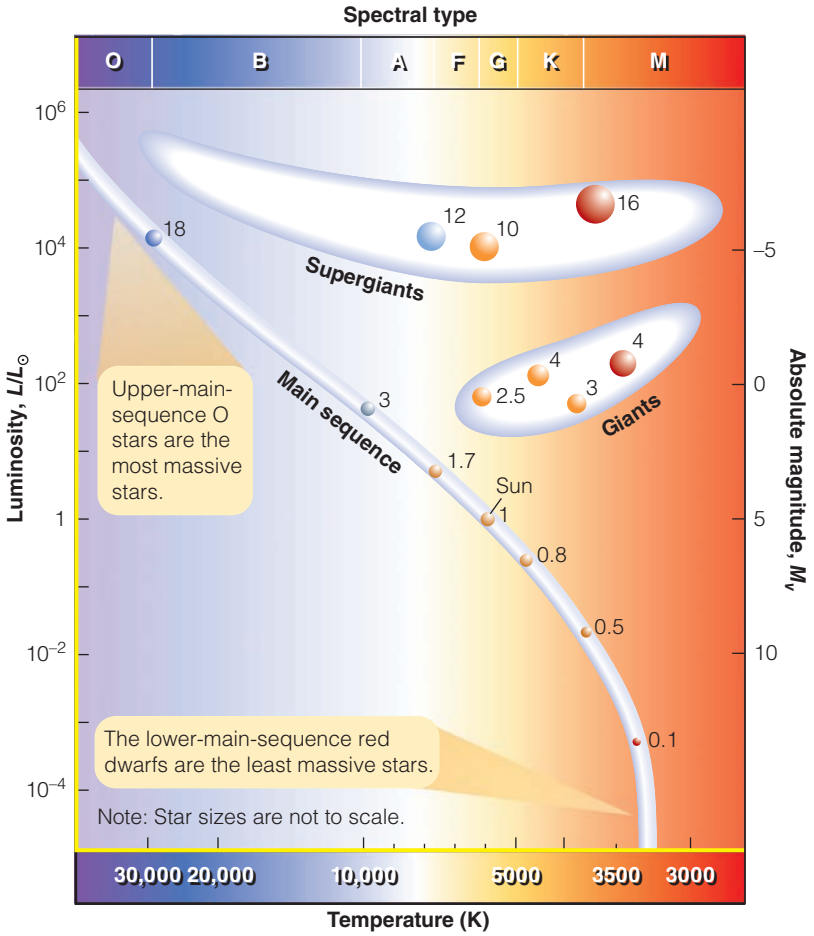
If you want to know what the average person thinks about a certain subject, you take a survey. If you want to know what the average star is like, you need to survey the stars. Such surveys reveal important relationships among the family of stars.

Not many decades ago, surveying large numbers of stars was an exhausting task, but modern computers have changed that. Specially designed computer-automated telescopes can make millions of observations per night, and other high-speed computers can compile those data into easy-to-use databases (Section 6-3a). Such surveys produce mountains of data that astronomers can "mine" while searching for relationships within the family of stars.

What could you learn about stars from a survey of the stars near the Sun? Astronomers have evidence that the Sun is in a typical place in the Universe. Therefore, a survey of nearby stars can reveal general characteristics of stars everywhere. Study **The Family of Stars** and notice three important points:

1. Making a good survey is difficult because you must be sure you get a good sample. If you don't survey enough stars, or if you miss some kinds of stars, your survey might produce biased results.
2. Luminous stars are rare. The most common types of stars are the low-luminosity red dwarfs.
3. A careful survey reveals that what you see in the sky is deceptive. Most of the bright stars in the sky are not typical stars but instead are highly luminous giants and supergiant stars that can be seen from great distances.

The night sky is a beautiful carpet of stars, and those stars have a tremendous variety.



▲**Figure 9-23** The masses of the plotted stars are labeled on this H–R diagram. Notice that the masses of main-sequence stars decrease systematically from upper left to lower right in the diagram but that masses of giants and supergiants are not arranged in any ordered pattern.

9-6b Mass, Luminosity, and Density

If you survey enough stars and plot the data in an H–R diagram, you can see the patterns that hint at how stars are born, how they age, and how they die.

If you label an H–R diagram with the masses of the plotted stars, as in **Figure 9-23**, you will discover that the main-sequence stars are ordered by mass. The most massive main-sequence stars are the hot stars, and as you run your eye along the main sequence, down and to the right in the diagram, you will find successively lower-mass stars until you reach the lowest-mass, coolest, faintest main-sequence stars, the red dwarfs.

In contrast, stars that are not on the main sequence are not ordered on the H–R diagram according to mass. Giant and supergiant stars are a jumble of different masses, although supergiants tend to be more massive than giants. Also, most white dwarfs have about the same mass, somewhere in the narrow range from 0.5 to about 1 solar mass.

9-6c The Mass–Luminosity Relation

The systematic ordering of mass along the main sequence means that the main-sequence stars follow a **mass–luminosity relation**—the more massive a star is, the more luminous it is

The Family of Stars

1 What is the most common type of star? What types of stars are rare? To answer those questions, you need to make a census of the stars. In doing so you collect information on their spectral classes, luminosity classes, and distances. Your survey of the family of stars produces some surprising demographic results.

1a You could survey the stars by observing every star within 62 pc of Earth. A sphere 62 pc in radius encloses a million cubic parsecs. Such a survey would tell you how many stars of each type are found within a volume of a million cubic parsecs.

2 Your survey faces two difficulties:

1. The most luminous stars are so rare you find only a few in your survey region. In fact, there is not even one O star within 62 pc of Earth.

2. Lower-main-sequence M stars—sometimes called red dwarfs—and white dwarfs are so faint they are hard to locate even when they are only a few parsecs from Earth. Finding every one of these stars in your survey sphere is a difficult task, but if you don't, your survey is incomplete.

Spectral Class Color Key

- O and B
- A
- F
- G
- K
- M

The star chart in the background of these two pages shows most of the constellation Canis Major; stars are represented as dots with colors assigned according to spectral class. The brightest stars in the sky tend to be the rare, highly luminous stars, which appear bright even though they are far away. Most stars are of very low luminosity, so nearby stars tend to be very faint red dwarfs.

62 pc

Earth

Red dwarf
15 pc

o^2 Canis Majoris
B3 Ia 790 pc

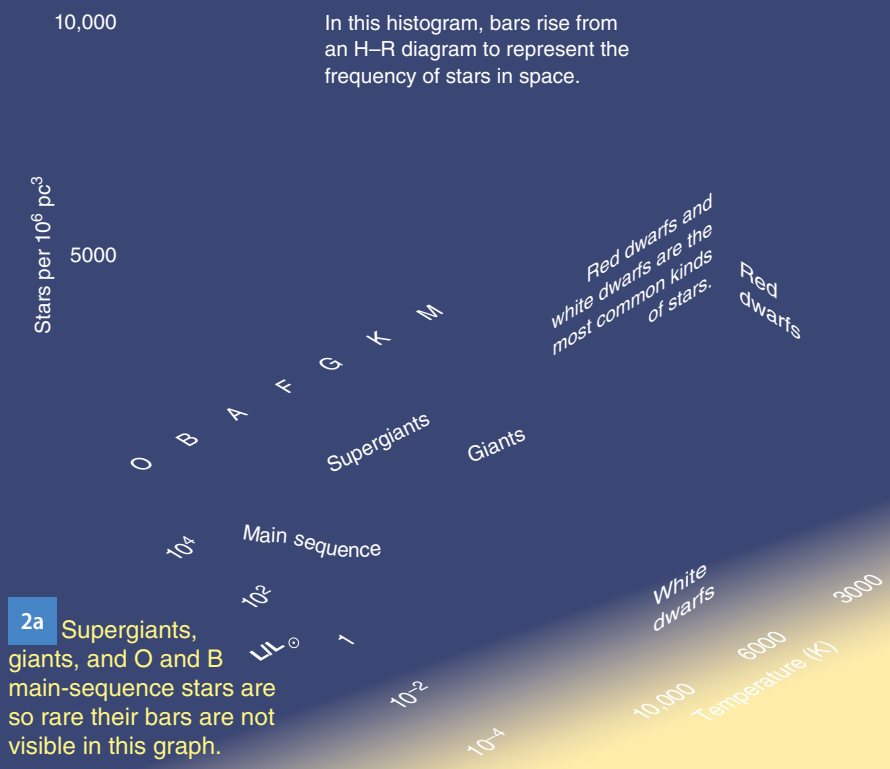
Red dwarf
17 pc

δ Canis Majoris
F8 Ia 550 pc

σ Canis Majoris
M0 Iab 370 pc

η Canis Majoris
B5 Ia 980 pc

ε Canis Majoris
B2 II 130 pc



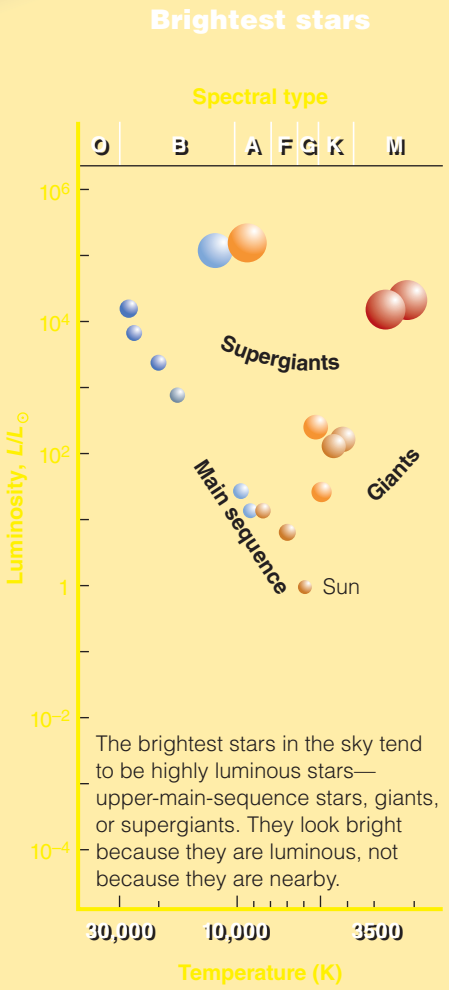
In this histogram, bars rise from an H–R diagram to represent the frequency of stars in space.

Sirius A (α Canis Majoris) is the brightest star in the sky. With a spectral type of A1 V, it is not a very luminous star. It looks bright because it is only 2.6 pc away.

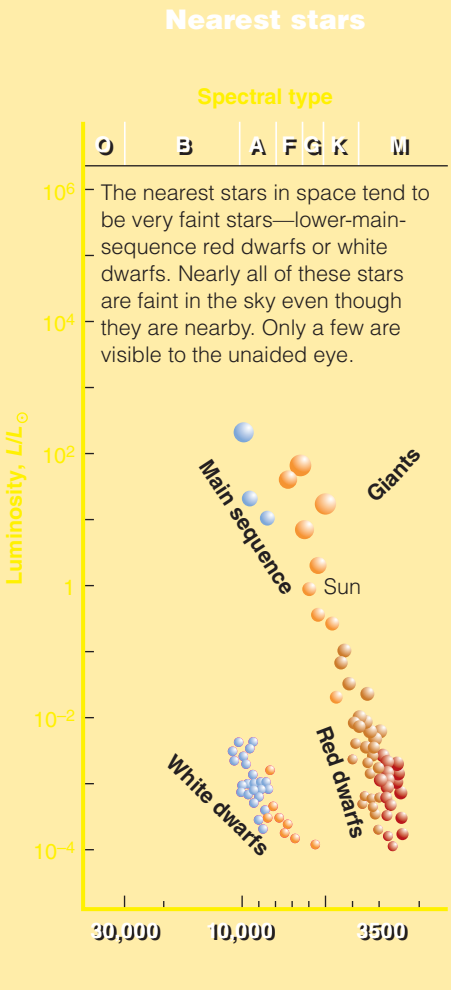
Sirius B is a white dwarf that orbits Sirius A. Although Sirius B is not very far away, it is much too faint to see with the unaided eye.

2a Supergiants, giants, and O and B main-sequence stars are so rare their bars are not visible in this graph.

3 Luminous stars are rare but are also easy to see even at great distances. Most stars are very low luminosity objects. Not a single white dwarf or red dwarf is bright enough to see with the unaided eye. See the H–R diagrams at right.



The brightest stars in the sky tend to be highly luminous stars—upper-main-sequence stars, giants, or supergiants. They look bright because they are luminous, not because they are nearby.



The nearest stars in space tend to be very faint stars—lower-main-sequence red dwarfs or white dwarfs. Nearly all of these stars are faint in the sky even though they are nearby. Only a few are visible to the unaided eye.

Basic Scientific Data

Where do large volumes of scientific data come from?

In one sense, science is the process by which scientists examine data and search for relationships, and it sometimes requires large amounts of data. For example, astronomers need to know the masses and luminosities of many stars before they can begin to understand the relationship between mass and luminosity.

Compiling basic data is one of the common forms of scientific work—a necessary first step toward scientific analysis and understanding. An archeologist may spend months or even years diving to the floor of the Mediterranean Sea to study an ancient Greek shipwreck. She will carefully measure the position of every wooden timber and bronze fitting. She will photograph and recover everything from broken pottery to tools and weapons. The care with which she records data on the site pays off when she begins her analysis. For every hour the archaeologist spends recovering an object, she may spend days or weeks in her office, a library, or a museum identifying and understanding the object. Why was there a Phoenician hammer on a Greek ship? What does that reveal about the economy of ancient Greece?

Finding, identifying, and understanding that ancient hammer contributes only a small bit of information, but the work of many scientists eventually builds a picture of how ancient Greeks saw their world. Finding the masses of the stars in one binary system does not tell an astronomer a great deal about nature. Over the years, however, many astronomers have added their results to the growing data file on stellar masses. Scientific data accumulate and can be analyzed by later generations of scientists.

Michael A. Seeds

Collecting mineral samples can be hard work but also fun.

(**Figure 9-24**). That relation can be expressed as a simple equation with which you can estimate the luminosity of a main-sequence star based on the star's mass. A main-sequence star's luminosity in units of the Sun's luminosity is approximately equal to its mass in solar masses raised to the power 3.5:

KEY EQUATION 9-5 The Stellar Mass-Luminosity Relation

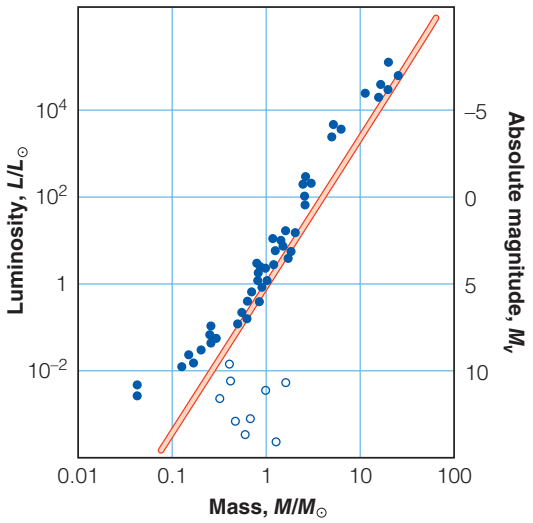
$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^{3.5}$$

That is the mathematical form of the mass–luminosity relation. ***Example:*** What is the luminosity of a star with four times the mass of the Sun? ***Solution:*** That star has $L/L_\odot = (M/M_\odot)^{3.5} = 4^{3.5} = 128\ L_\odot$. The same relation is also expressed by the red line in Figure 9-24, and you can see that it is only an approximation to the data. Nevertheless, the relation applies fairly well to most main-sequence stars over a wide range of stellar masses.

Notice the large range of luminosity in Figure 9-24. The figure covers a portion of the total observed range of stellar masses, which extends from about 0.1 solar mass to about 100 solar masses—a factor of 1000. But, the range of luminosities extends from about a few times $10^{-5}\ L_\odot$ to more than $10^6\ L_\odot$—a factor of a hundred billion (10^{11}). Clearly, a moderate difference in mass causes a large difference in luminosity.

Although giants, supergiants, white dwarfs, and brown dwarfs do not follow simple mass–luminosity relationships, the link between mass and luminosity for main-sequence stars is fundamental to astronomy. In the next chapters, the mass–luminosity relation will help you understand how stars generate their energy.

The density of stars reveals another pattern in the H–R diagram. The average density of a star is its mass divided by its volume. As you will learn in the next chapter, stars are not uniform in density but are most dense at their centers and least dense near their surface. The center of the Sun, for instance, is about 150 times as dense as water (15 times denser than lead); its density near the visible surface is less than 1/3000 times as dense as Earth's atmosphere at sea level. The Sun's average density is approximately 1.4 g/cm^3—just a little denser than water, intermediate between its central and surface densities.



▲**Figure 9-24** The mass–luminosity relation shows that the more massive a main-sequence star is, the more luminous it is. The open circles represent white dwarfs, which do not obey the relation. The red line represents the equation $L = M^{3.5}$.

Main-sequence stars have average densities similar to the Sun's, but giant stars have low average densities, ranging from 0.01 to 0.1 g/cm^3. The enormous supergiants have still lower densities, ranging from 10^{-7} to 10^{-3} g/cm^3. These densities are thinner than the air you breathe, and if you could insulate yourself from the heat, you could fly through these stars. Only near their centers would you be in any danger because there the material is very dense—millions of times denser than water.

The white dwarfs have masses about equal to the Sun's but are very small, only about the size of Earth. That means the matter is compressed to enormous densities. On Earth, a teaspoonful (5 cc) of this material would weigh more than 5 tons.

Density divides stars into three groups. Most stars are main-sequence stars with densities like the Sun's. Giants and supergiants are very low-density stars, and white dwarfs are high-density stars. You will see in later chapters that these groups of densities represent different stages in the evolution of stars.

Practicing SCIENCE

What kind of stars do you mostly find if you look at a few of the brightest stars in the sky? This question demonstrates how careful a scientist must be in interpreting even seemingly simple observations.

When you look at the night sky, the brightest stars are mostly giants and supergiants. For example, look at the Concept Art: The Family of Stars. Most of the bright stars in Canis Major are supergiants. Sirius, one of the Famous Stars, is an exception. It is the brightest star in the sky, but it is only a main-sequence star that looks bright because it is very nearby, not because it is very luminous. In general, supergiant and giant stars are so luminous that they stand out and look bright, even though they are not nearby. When you look at a bright star in the sky, you are probably looking at a highly luminous star—a supergiant or a giant. You can check this argument by consulting the tables of the brightest and nearest stars in Appendix A.

Now turn the question around: ***What kinds of stars do you mostly find if you catalog the stars nearest to the Sun?***

WHAT ARE WE?

Medium

We humans are medium-sized creatures, and we experience medium-sized things. You can see trees and flowers and small insects, but you cannot see the beauty of the microscopic world without ingenious instruments and special methods. Similarly, you can sense the grandeur of a mountain range, but larger objects, such as the Milky Way Galaxy, are too big for our medium-scale senses. You have to use your ingenuity and imagination to experience the truth of such large objects. That is what science does for us. We live between the microscopic world and the astronomical world, and science enriches our lives by revealing the parts of the Universe beyond our daily experience.

Experience is fun, but it is very limited. You may enjoy a flower by admiring its color and shape and by smelling its fragrance. But the flower is more wonderful than your direct experience can reveal. To truly appreciate the flower, you need to understand it—how complex it truly is, how it serves its plant, and how the plant evolved to make such a beautiful blossom.

Humans have a natural drive to understand as well as experience. You have experienced the stars in the night sky, and now you are beginning to understand them as objects ranging from hot blue O stars to cool red M dwarfs. It is natural for you to wonder why these stars are so different. As you explore that story in the following chapters, you will discover that, although you have medium-scale senses, you can understand the stars.

Study and Review

SUMMARY

9-1 Star Distances

- Only by first knowing the distance to a star can you find a star's other properties. To calculate the distance of a star, astronomers use a star's parallax, or the apparent change in its position as a result of a change in the location of the observer.
- Astronomers can measure the distances to nearer stars by observing their **stellar parallaxes (*p*),** using the parallax equation:

$$d\,(\text{pc}) = \frac{1}{p\,(\text{arc seconds})} \qquad \text{(Eq. 9-1)}$$

- Stellar distances are commonly expressed in units of **parsecs (pc)**. One parsec is 2.06×10^5 AU and is the distance to an object having a stellar parallax of 1 arc second as measured from Earth. One parsec equals 3.26 light-years. **Proper motions,** or the rate at which a star moves across the sky, can give you a clue to distance because nearby stars tend to have large proper motions.

9-2 Apparent Brightness, Intrinsic Brightness, and Luminosity

- The amount of light received from a star, the light **flux**, is related to its distance by the inverse square law. Once you know the distance to a star and its apparent magnitude, you can use the **magnitude–distance formula**:

$$m_v - M_v = -5 + 5\log(d) \qquad \text{(Eq. 9-2)}$$

to find its **intrinsic brightness**, which is expressed as the star's **absolute visual magnitude (M_v)**, the apparent magnitude the star would have if the star were located at a distance of 10 pc.

- To find the energy output of a star, astronomers must make a correction for the light arriving with wavelengths that are not visible to convert absolute visual magnitude into **luminosity (*L*)**, which is the total energy radiated by the star's surface in 1 second.

9-3 Stellar Spectra

- The strength of spectral lines depends significantly on the temperature of the star. For example, in cool stars, the hydrogen Balmer lines are weak because atoms are not excited out of the ground state. In hot stars, the Balmer lines are weak because atoms are excited to higher orbits or are ionized. Only at intermediate temperatures, around 10,000 K, are the Balmer lines strong.
- A star's temperature **spectral type** (or class) is determined by which absorption lines are visible in its spectrum. The **spectral sequence**, OBAFGKM, is important because it is a temperature sequence. By classifying a star by spectral type, the astronomer learns the temperature of the star's surface.
- Long after the spectral sequence was created, astronomers found the **L dwarfs**, **T dwarfs,** and **Y dwarfs** with temperatures even cooler than the M stars. These are examples of **brown dwarfs**, objects smaller than stars but larger than planets.

9-4 Star Sizes

- A star's luminosity can be calculated in proportion to the Sun with this formula:

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \left(\frac{T}{T_\odot}\right)^4 \qquad \text{(Eq. 9.3)}$$

- The **Hertzsprung–Russell (H–R) diagram** is a plot of stellar luminosity versus surface temperature. It is an important diagram in astronomy because it sorts the stars into categories by size.
- Roughly 90 percent of stars, including the Sun, fall on the **main sequence** of the H–R diagram, with the more massive stars being hotter and more luminous. The **giants** and **supergiants** are much larger than main-sequence stars and lie above the main sequence in the diagram. **Red dwarfs** lie at the bottom end of the main sequence. Some of the **white dwarfs** are very hot stars, but they fall below the main sequence because they are so small.
- Observations of a few stars made with interferometers confirm the sizes implied by the H–R diagram.
- The large sizes of the giants and supergiants mean their atmospheres have low densities and their spectra have sharper spectral lines than the spectra of main-sequence stars. In fact, it is possible to assign stars to **luminosity classes** by the widths of their spectral lines. Class V stars are main-sequence stars. Class III stars, giants, have sharper lines, and Class I stars, supergiants, have extremely sharp spectral lines.
- Astronomers can use the locations of the spectral type and luminosity class in the H–R diagram to estimate the distances to a star using a technique called **spectroscopic parallax**.

9-5 Star Masses—Binary Stars

- The only direct way to find the masses of stars is by studying **binary stars**, in which two stars orbit their common center of mass. Astronomers find the masses of the stars by observing the period and sizes of their orbits and then using Newton's version of Kepler's third law:

$$M_A + M_B = \frac{a^3}{P^2} \qquad \text{(Eq. 9-4)}$$

- In a **visual binary**, both stars are visible and the orbits can be mapped, but in a **spectroscopic binary**, the stars are so close together they look like a single point of light, and the orbits can't be observed directly. In an **eclipsing binary**, the orbits are nearly edge-on, and the stars cross in front of each other. The resulting brightness changes in the **light curve** can reveal the diameters of the stars as well as their masses.

9-6 A Census of the Stars

- A survey in the neighborhood of the Sun shows that lower-main-sequence stars, less luminous than the Sun, are the most common type. White dwarfs are also fairly common, although they are faint and hard to find. Giants and supergiants are rare.
- The **mass–luminosity relation** expresses the fact that the more massive a main-sequence star is, the more luminous the star is:

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^{3.5} \qquad \text{(Eq. 9-5)}$$

  The most massive main-sequence stars are type O at the upper left of the H–R diagram, and the least massive are type M at the lower right of the diagram. Giant and supergiant stars do not follow simple mass–luminosity relations, and neither do white dwarfs.
- Given the mass and diameter of a star, you can find its average density. Main-sequence stars have about the same density as the Sun, but giants and supergiants are very low-density stars. Some have much lower average density than Earth's air. The white dwarfs, lying below the main sequence, are tremendously dense.

REVIEW QUESTIONS

1. Why are Earth-based parallax measurements limited to the nearest stars?
2. Why was the *Hipparcos* satellite able to make more accurate parallax measurements than ground-based telescopes?
3. Star A and Star B have measured stellar parallax of 1.0 arc second and 0.75 arc second, respectively. Which star is closer? How do you know?
4. What do the words *absolute* and *visual* mean in the definition of *absolute visual magnitude?* Similarly, what do the words *apparent* and *visual* mean in the definition of *apparent visual magnitude?* How are the two related?
5. If a star's apparent magnitude is equal to its absolute magnitude, what must be the star's distance?
6. A star is observed to have no proper motion. What are some possible explanations?
7. What does luminosity measure that is different from what absolute visual magnitude measures?
8. Why are hydrogen Balmer lines strong in the spectra of medium-temperature stars and weak in the spectra of hot and cool stars? Is the Sun a cool, medium, or hot star, based on the Balmer thermometer?
9. Why are TiO (titanium oxide) features visible in the spectra of only the coolest stars?
10. You observe strong FeH (iron hydride) lines but weak CH_4 (methane) lines in the spectra of a celestial object. What spectral class would you assign to the object based on the strength of these lines? What kind of object is that?
11. Explain the interrelationships among Table 9-1, Figure 9-5c, Figure 9-6, and Figure 9-7.
12. Why does the luminosity of a star depend on both its radius and its temperature?

13. If luminosity depends on radius and temperature, what does flux depend on?
14. Star A's luminosity is four times greater than Star B's. Both stars have the same surface temperature. Which star is larger? If both stars are on the main sequence, which star is more massive?
15. The difference in surface temperature of two stars is 5500 K, but both stars put out the same amount of power at their surfaces. One star is ten times larger than another star. One star is an M dwarf. What kind of star is the other? Using Figure 9-12, identify the two stars that best fit these data.
16. How can you be sure that supergiants really are larger than main-sequence stars?
17. A star is listed in the catalog as being a type G2 V. What is the star's spectral class? What is the star's luminosity class? What type of star is it? What is the peak color of the star?
18. What evidence shows that white dwarfs must be very small?
19. A star has a spectral class M. What type of star could this be?
20. A star is known to be much larger than a main-sequence star and is blue in color. What kind of star is it?
21. What observations would you make to classify a star according to its luminosity? Why does that method work?
22. Why does the orbital period of a binary star depend on the binary's mass?
23. A star seems to have no measurable stellar parallax. Can you still find the distance to the star? If so, how? If not, why not?
24. What observations would you make to study an eclipsing binary star, and what would those measurements tell you about the component stars?
25. Why don't astronomers know the inclination of a spectroscopic binary? How do they know the inclination of an eclipsing binary?
26. You observe a star system for many weeks. The light curve does not show significant changes in brightness, but a series of spectra show a spectral feature that cyclically appears as one absorption line, then two, then one again, and so on. Is this a binary system? If so, what type? What can you determine about the objects in the system? How?
27. How do the masses of stars along the main sequence illustrate the mass–luminosity relation?
28. Why is it difficult to find out how common the most luminous stars are? The least luminous stars?
29. What is the most common type of star?
30. Which is the most common non–main-sequence star type?
31. If you look only at the brightest stars in the night sky, what type of star are you likely to be observing? Why?
32. **How Do We Know?** What is the missing link in the chain of inference leading from observations of spectroscopic binaries to the masses of stars?
33. **How Do We Know?** In what way are basic scientific data cumulative, and how do accumulations of such data help later scientists?

ACTIVE INQUIRY QUESTIONS

1. If someone asked you to compile a list of the nearest stars to the Sun based on your own observations, how would you select your sample, what measurements would you make, and how would you analyze the measurements to detect nearby stars?
2. Can you think of classification systems used to simplify what would otherwise be complex measurements? Consider foods, movies, cars, grades, and clothes.
3. Your marketing agency has been hired to rebrand the Sun as something other than an average star. What slogan can you come up with?

PROBLEMS

1. If a star has a parallax of 0.050 arc second, what is its distance in pc? In ly? In AU?
2. Celestial object A is located 5 ly from Earth. Celestial object B is located 5 pc from Earth. Celestial object C is located 5 AU from Earth. Convert the distances of objects B and C to light-years and rank the objects from farthest to closest from Earth. Give an example of which kind of object could be located at each of these distances.
3. Star A is located five times farther from Earth than Star B. Star B has a 0.020 arc second stellar parallax. How far from Earth is Star A, and what is its stellar parallax? Could ground-based telescopes measure Star A's parallax?
4. If a star has a parallax of 0.016 arc second and an apparent magnitude of 6.0, how far away is it, and what is its absolute magnitude?
5. Complete the following table:

M (mag)	M_v (mag)	d (pc)	p (arc seconds)
____	7	10	____
11	____	1000	____
____	22	____	0.025
4	____	____	0.040

6. Star A is located four times farther from Earth than Star B, but both have the same apparent visual magnitude of 1 mag. Which star is intrinsically brighter and by how much?
7. Determine the spectral types and temperatures of the following stars based on the appearances of their spectra. Refer to Figure 9-5c.
   a. Medium-strength hydrogen Balmer lines, strong helium lines
   b. Medium-strength hydrogen Balmer lines, weak ionized-calcium lines
   c. Strong TiO bands
   d. Weak hydrogen Balmer lines, strong ionized-calcium lines
8. You find the spectra of a celestial object and classify it as spectral type T4. In which band of the electromagnetic spectrum would this object mostly emit? What color would you see this celestial object? Is this object a star? How do you know? (*Hints:* Use Wien's law (Eq. 7-2) and Figure 6-3.)
9. If a main-sequence star has a luminosity of 400 $L_\odot$, what is its spectral type? (*Hint:* See Figure 9-12.)
10. The apparent visual magnitude of Arcturus is −0.04. Use the method of spectroscopic parallax to calculate the approximate distance to Arcturus in units of pc. What is the luminosity class of Arcturus? (*Hint:* Examine Figure 9-12.)
11. If a star has an apparent magnitude equal to its absolute magnitude, how far away is it in parsecs? In light-years?
12. An O8 V star has an apparent visual magnitude of 11. Use the method of spectroscopic parallax to estimate the distance to the star. (*Hints:* Refer to one of the H–R diagrams in the chapter, and use the magnitude–distance formula, Eq. 9-2.)
13. Assume that an O main-sequence star and a B main-sequence star have the same radius. What would be the ratio of their luminosities? (*Hint:* Use the Stefan-Boltzmann law, Eq. 7-1.) (*Note:* Necessary data are given in Table 9-1.)
14. At the position of Earth, the total flux of sunlight at all wavelengths is 1370 W/m^2. Find the luminosity of the Sun. Make your calculation in two steps. First, use $4\pi R^2$ to calculate the surface area in square meters of a sphere surrounding the Sun with a radius of 1 AU. Second, multiply by the solar constant to find the total solar energy passing through the sphere in 1 second. That is the luminosity of the

Sun in units of watts. Compare your result with that in **Celestial Profile: The Sun**, Chapter 8.

15. In the following table, which star is brightest in apparent magnitude? Most luminous in absolute magnitude? Largest in diameter? Farthest away?

Star	Spectral Type	m_V (mag)
a	G2 V	5
b	B1 V	8
c	G2 Ib	10
d	M5 III	19
e	White dwarf	15

16. What is the total mass of a visual binary system if the average separation of the stars is 8.0 AU and their orbital period is 20 years?
17. In a binary star system, the average separation between the stars is 5.0 AU, and their orbital period is 5.0 years. What is the sum of their two masses? The average distance of Star A from the center of mass is four times that of Star B. What are their individual masses?
18. If an eclipsing binary has a period of 32 days, and the two components are in circular orbits around their center of mass with speeds respectively of 154 km/s and 77 km/s, what are the circumferences of the two orbits? What is the separation between the two stars? The total mass of the system? The mass of each star?
19. Estimate the luminosity of a 3-solar-mass main-sequence star; of a 9-solar-mass main-sequence star. Can you easily estimate the luminosity of a 3-solar-mass red giant star? Why or why not?
20. One main-sequence star is three times more massive than another main-sequence star. How much more power does the more massive star put out at its surface than the other? If the less massive star's $M_V = 5$, which are the two stars that best fit this data according to Figure 9-12 and Figure 9-23?

SENSE OF PROPORTION

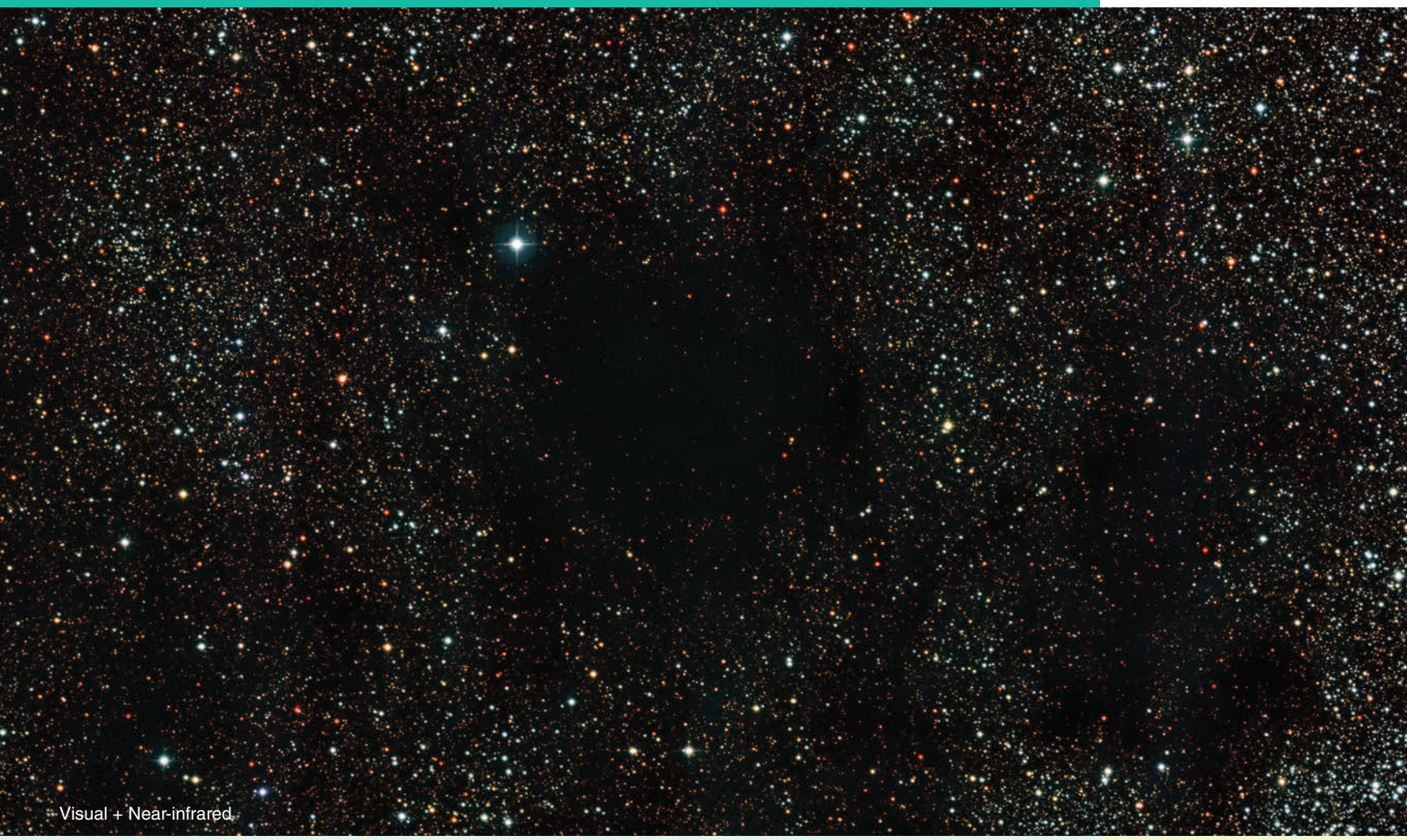
1. The largest red supergiant stars have about 1000 times the diameter of the Sun. If you represent the Sun with a golf ball (diameter 1.68 in.), how big an object would represent a red supergiant? Search online to find a familiar object that is about that size.
2. White dwarfs have about 1/100 the diameter of the Sun. If you represent the Sun with a golf ball (diameter 1.68 in.), how small a sphere would represent a white dwarf? Search online to find a familiar object that is about that size.
3. Arrange the following star classes going from the most common type to the least common type: O main sequence, white dwarf, M main-sequence star.

LEARNING TO LOOK

1. Look at the image on the opening page of this chapter. Some white stars are shown to be very large circles with spikes. Why are these stars apparently very large? What causes the spikes?
2. Look at Figure 9-2. A star appears to have moved 1.5 arc seconds in the sky from where it was located 6 months ago. Using the conventional definition of parallax, assuming a baseline of 1 AU, what is the parallax angle *p*?
3. Look at Figure 9-9. Why is the lava nearest the source brighter and yellower than the lava that is farther away?
4. If all of the stars in this photo are members of the same star cluster, then they all are about the same distance from Earth. Then why are three of the brightest much redder than the rest? What kind of star are they?

NASA

10 The Interstellar Medium



ESO

▲ An image of part of the huge cloud of dust and gas known as the Coalsack Nebula. The dust in the nebula absorbs and scatters almost all visible light from background stars.

Having begun your study of the Sun and other stars, it is now time to focus on the thin gas and dust that drift through space between the stars, known as the interstellar medium. This chapter will help you answer three important questions:

- How do astronomers study the interstellar medium?
- What kinds of material make up the interstellar medium?
- How does the interstellar medium interact with stars?

In the next chapter you will learn that the interstellar medium is the raw material from which stars form. The gas and dust between the stars is the starting point for the life story of the stars, including our Sun. The next four chapters will trace the birth, life, and death of stars.

. . . when he shall die,
Take him and cut him out in little stars,
And he will make the face of heaven so fine
That all the world will be in love with night
And pay no worship to the garish sun.

—WILLIAM SHAKESPEARE, *ROMEO AND JULIET*, 3.2.21

JULIET LOVED ROMEO SO MUCH she compared him to the beauty of the stars. Had she known what was between the stars, she might have compared him to that instead. The material between the stars, called the **interstellar medium (ISM)**, is mostly invisible to human eyes, but when observed at infrared wavelengths it is strikingly beautiful, and its densest clouds are the nurseries where new stars are born. If there is beauty in vast extent and sweeping power, then the ISM could steal worship from the garish suns.

When people use the phrase "the vacuum of space," they reveal a **Common Misconception** that space is empty. Space is not empty; in fact, you are about to discover that the ISM is full, complex, and active (**Figure 10-1**).

10-1 Studying the Interstellar Medium

There is nothing special about visible light except that human eyes can see it. You will need to use all parts of the electromagnetic spectrum to fully understand the ISM.

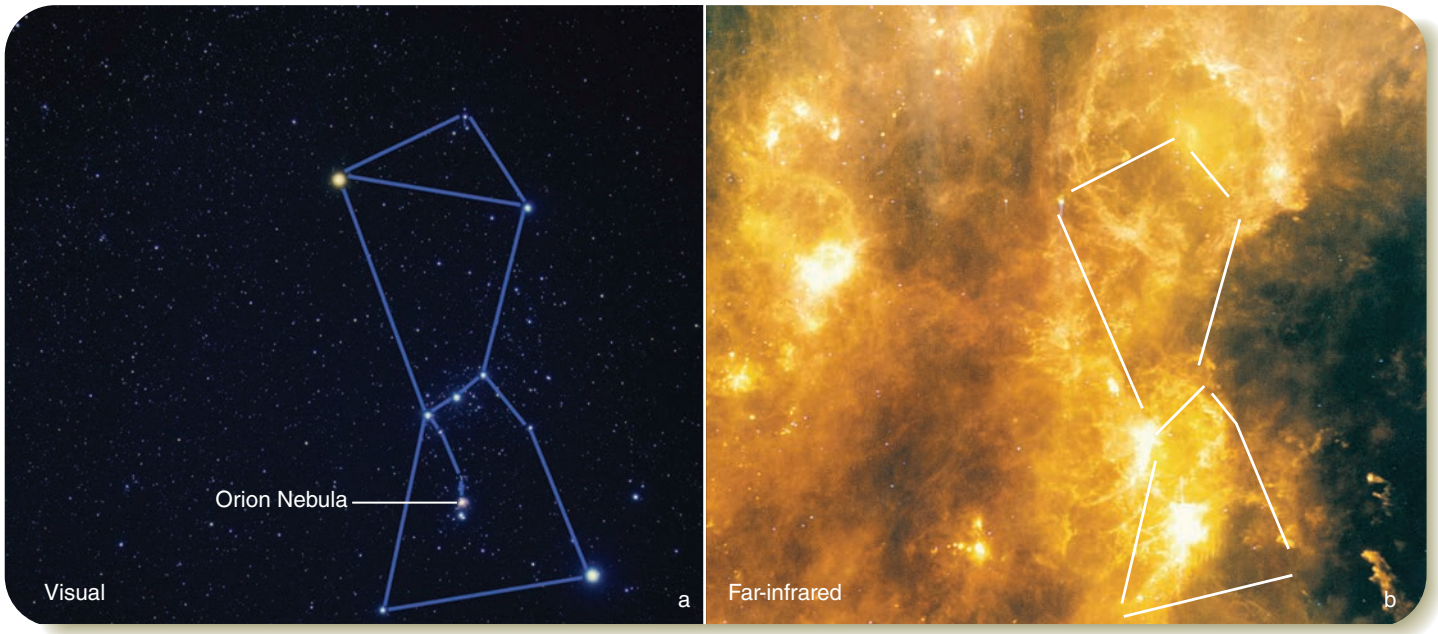
10-1a Nebulae

On a cold, clear winter night, Orion hangs high in the southern sky, a large constellation composed of brilliant stars. If you carefully observe Orion's sword, you will see that one of the stars is a hazy cloud (look back to Figure 2-4). A small telescope reveals even more such clouds of gas and dust scattered around the sky. Astronomers refer to these clouds as **nebulae** (singular, nebula), from the Latin word for cloud or fog. They are visible evidence of the ISM.

Study **Three Kinds of Nebulae** and notice three important points and four new terms:

1. *H II regions* (H plus Roman numeral II, pronounced *aitch-two*) are clouds of gas ionized by the ultraviolet radiation from nearby hot stars. They are often called *emission nebulae* because the ionized hydrogen produces visible wavelength photons that make the regions glow with a characteristic pink color.
2. If nearby stars are not hot enough to ionize the hydrogen in a nebula, it may still be visible to human eyes as a *reflection nebula,* produced when light is scattered by tiny dust specks mixed in with the gas.
3. At visual wavelengths, *dark nebulae* are visible where dense clouds of gas and dust are silhouetted against background regions filled with stars or bright nebulae.

The different kinds of nebulae that decorate the sky are really similar objects seen in different ways, much like the various types of clouds in Earth's atmosphere. On a sunny day, a distant cloud looks white, but overhead it blocks sunlight and looks dark. If you could see the same cloud with infrared eyes, you would see it glowing brightly with



Panel a: A.Fujii; Panel b: NASA/IRAS/IPAC, Courtesy D.Levine

▲**Figure 10-1** (a) At visual wavelengths, the spaces between the stars in and around the constellation Orion seem empty. (b) An infrared image reveals the swirling clouds of the ISM.

blackbody radiation. Different kinds of nebulae are regions of the ISM with different relationships to starlight and to you, the observer.

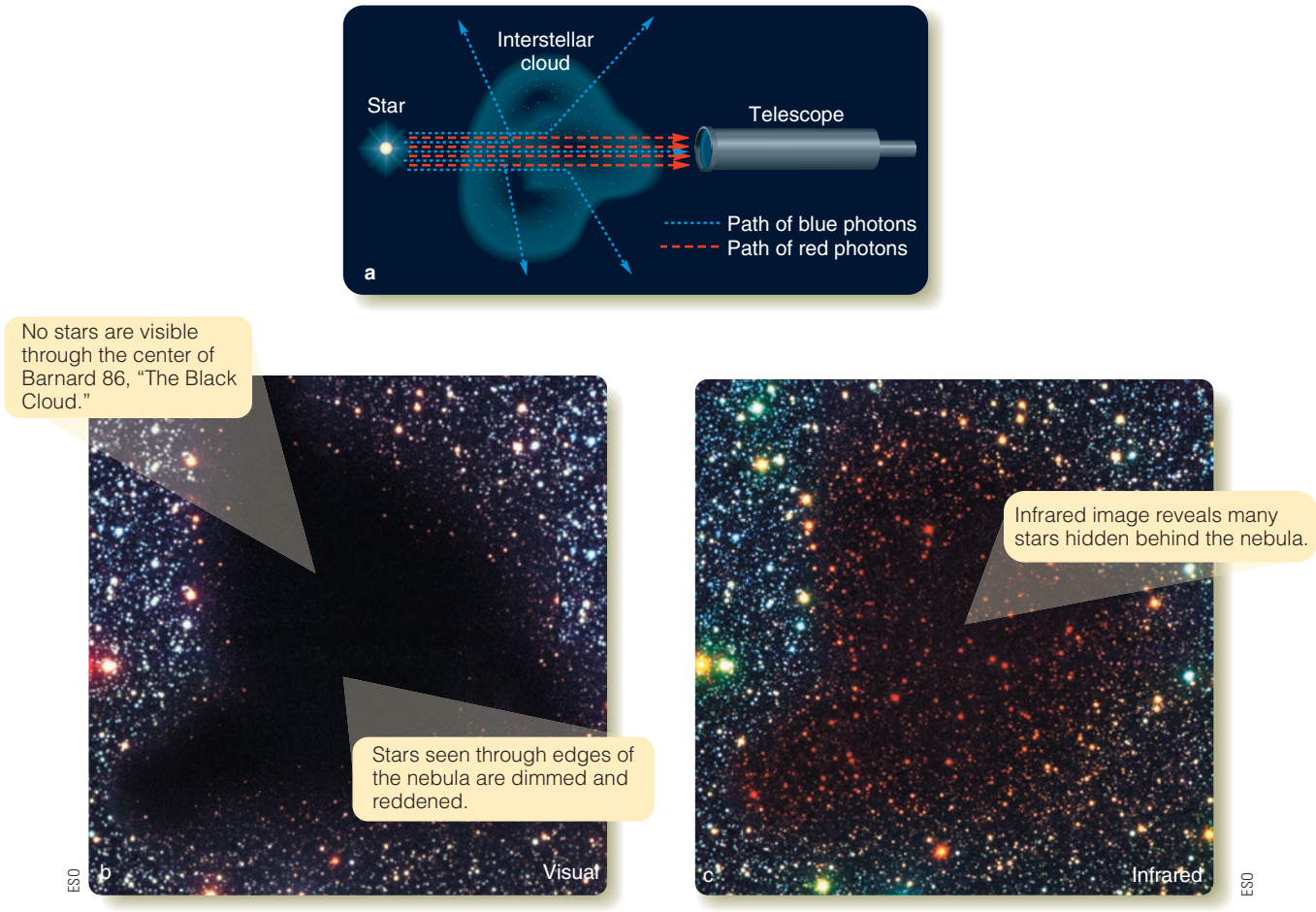
Detailed analysis of the spectra of nebulae reveals their composition. About 70 percent of the mass in the ISM is hydrogen, 28 percent is helium, and 2 percent elements heavier than helium. Note that this is about the same composition as the Sun and other stars (look back to Table 8-1). The density of the ISM is very low, typically only 0.1 to 1 atom per cubic centimeter, although it can be much higher in dense clouds. Even in those clouds, the gas is 10^{15} times less dense than the air you breathe.

Almost all of the ISM is in the form of gas, but about 1 percent of its mass is made of dust particles called **interstellar dust**. The dust particles are composed of carbon, silicates, iron, and, in some locations, ice plus organic compounds. (An organic compound is made up of molecules that have a carbon chain structure, but does not have to originate with living things.) ISM dust particles are almost all less than 1 micron (only 10^{-6} meters or 1000 nm) in diameter.

10-1b Extinction and Reddening

One way to study the ISM is to examine its effect on the starlight that passes through it. The dust in the ISM makes distant stars appear fainter than they would if space were perfectly transparent. This phenomenon is called **interstellar extinction**, and in the neighborhood of the Sun it amounts to about 1 magnitude per thousand parsecs (pc) at visual wavelengths. That is, a star 1000 pc from Earth will look about 1 magnitude fainter to human eyes than it would if space were perfectly empty. If it is 2000 pc away, it will look about 2 magnitudes dimmer, and so on. This is a strong effect at visual wavelengths and shows that the ISM is not confined to a few nebulae scattered here and there; it is everywhere in space.

The ISM dust also affects the colors of stars. As you saw previously, the dust particles are small, with diameters equal to or smaller than the wavelengths of light, so they scatter photons with shorter wavelengths more efficiently than those with longer wavelengths. That is the reason why extinction is more severe at shorter wavelengths, which makes stars look redder than they should be. Some hot but distant O stars, for instance, actually look red and not blue. This effect is called **interstellar reddening** (Figure 10-2).



▲**Figure 10-2** (a) Interstellar reddening makes stars seen through a cloud of gas and dust look redder than they should because shorter wavelengths are more easily scattered. (b) No stars are detected at visual wavelengths through a dense cloud except near its edges. (c) At longer near-infrared wavelengths, many stars can be detected behind the cloud, although their light is reddened.

Three Kinds of Nebulae

1 **Emission nebulae** are produced when a hot star excites the gas near it to produce an emission spectrum. The star must be hotter than about spectral type B1 (25,000 K). Cooler stars do not emit enough ultraviolet radiation to ionize the gas. Emission nebulae have a distinctive pink color produced by the blending of the red, blue, and violet hydrogen Balmer lines. Emission nebulae are also called **H II regions**, following the convention of using Roman numerals to indicate an atom's ionization state. Thus, H I means neutral hydrogen; H II is ionized.

In an H II region, the ionized nuclei and free electrons are mixed. When a nucleus captures an electron, the electron falls down through the atomic energy levels, emitting photons at specific wavelengths. Spectra indicate that the nebulae have compositions much like that of the Sun—mostly hydrogen. Emission nebulae have densities of 100 to 1000 atoms per cubic centimeter, thinner than the best vacuums produced in laboratories on Earth.

Visual

Anglo-Australian Observatory/David Malin Images

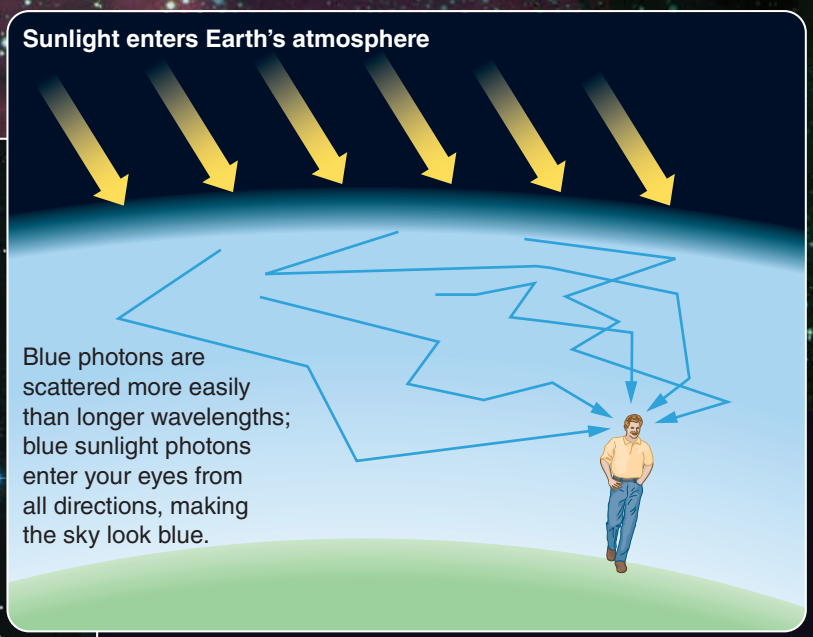
2 A **reflection nebula** such as the one shown in the photo below is produced when starlight scatters from the dust in a nebula. Consequently, the spectrum of a reflection nebula is just the reflected spectrum of the starlight, atmospheric absorption lines included. Gas is surely present in a reflection nebula, but it is not excited enough to emit photons.

Reflection nebulae NGC 1973, 1975, and 1977 are just north of the Orion Nebula. The pink glow from ionized hydrogen is an emission nebula deep inside the complex of reflection nebulae.

European Southern Observatory

Visual

2a Reflection nebulae look blue for the same reason the sky looks blue. Short wavelengths scatter more easily than long wavelengths. See sketch below.


Sunlight enters Earth's atmosphere
Blue photons are scattered more easily than longer wavelengths; blue sunlight photons enter your eyes from all directions, making the sky look blue.

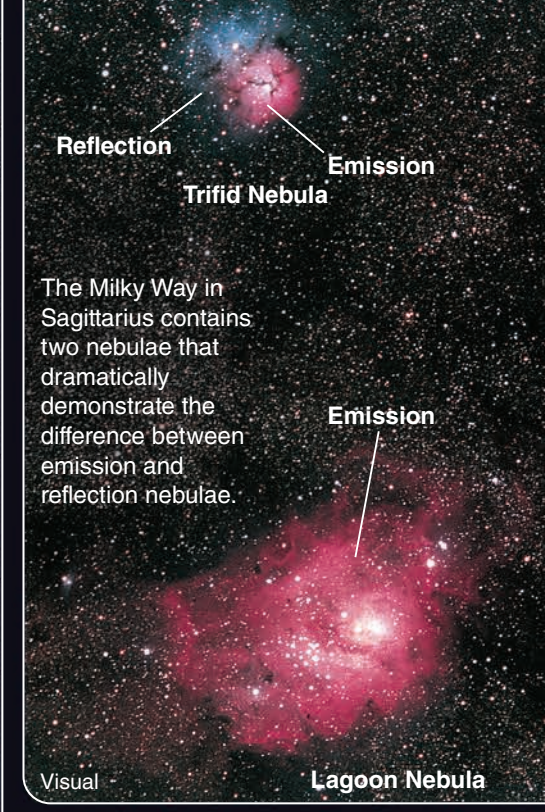
Anglo-Australian Observatory/David Malin Images

2b The blue color of reflection nebulae at left shows that the dust particles must be very small to preferentially scatter the blue photons. Interstellar dust grains must have typical

This particular re

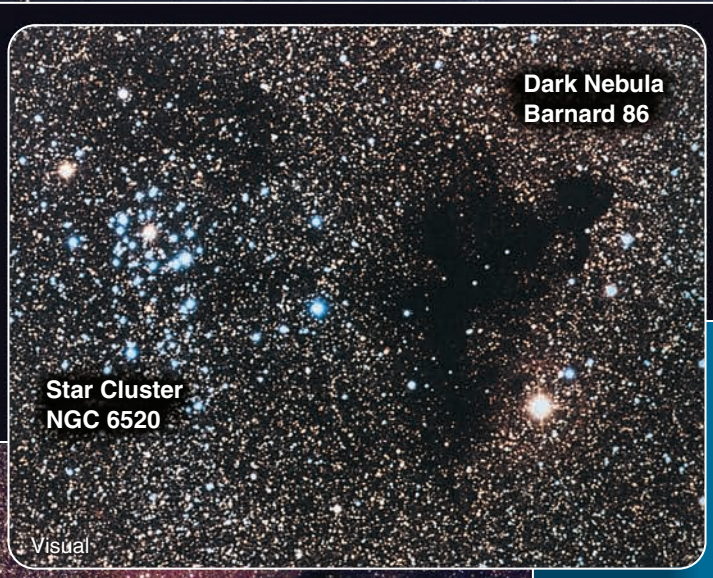
2c The hottest stars in the Pleiades star cluster are spectral type B6, not hot enough to ionize hydrogen in the ISM. Instead, the brightest stars produce a reflection nebula as their light is scattered from interstellar dust in a cloud that the cluster happens to be passing through.

Visual



Daniel Good

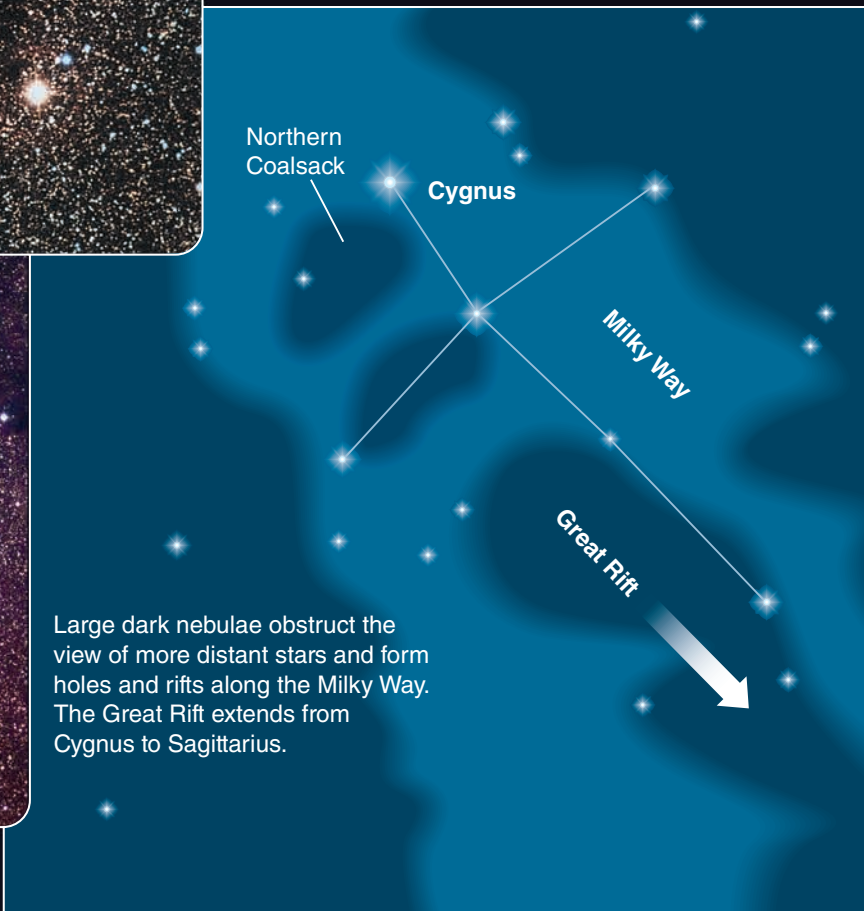
NASA/ESA/STScI/AURA/NSF/Caltech, Palomar Observatory/D. Soderblom, E. Nelan, F. Benedict, B. Arthur



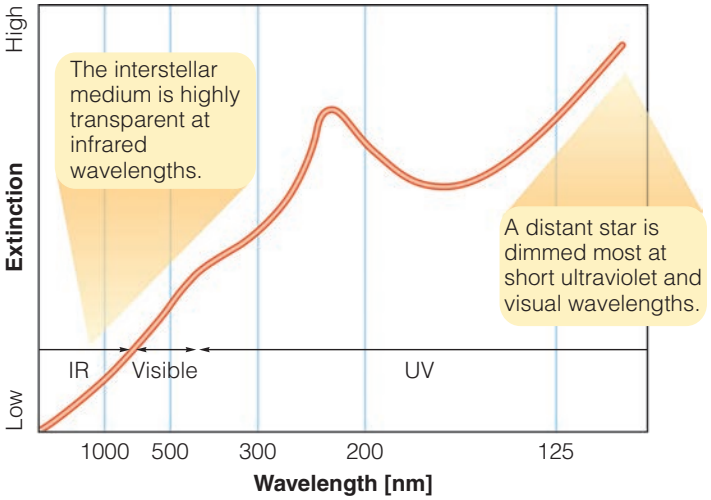
Anglo-Australian Observatory/ David Malin Images

3 **Dark nebulae** are clouds of gas and dust dense enough to obstruct the view of more distant stars. Their shapes, as shown at left and lower left, are generally irregular, suggesting that, in addition to the effects of nearby stars, there are breezes and currents pushing through the ISM.

Anglo-Australian Observatory/ David Malin Images



Large dark nebulae obstruct the view of more distant stars and form holes and rifts along the Milky Way. The Great Rift extends from Cygnus to Sagittarius.



◀**Figure 10-3** Interstellar extinction, the dimming of starlight by dust between the stars, depends strongly on wavelength. Infrared radiation is only slightly affected, but ultraviolet light is strongly affected. (Note that wavelengths increase to the left in this plot.) The relatively strong extinction at about 220 nm is caused by a spectral feature of a form of carbon dust in the ISM.

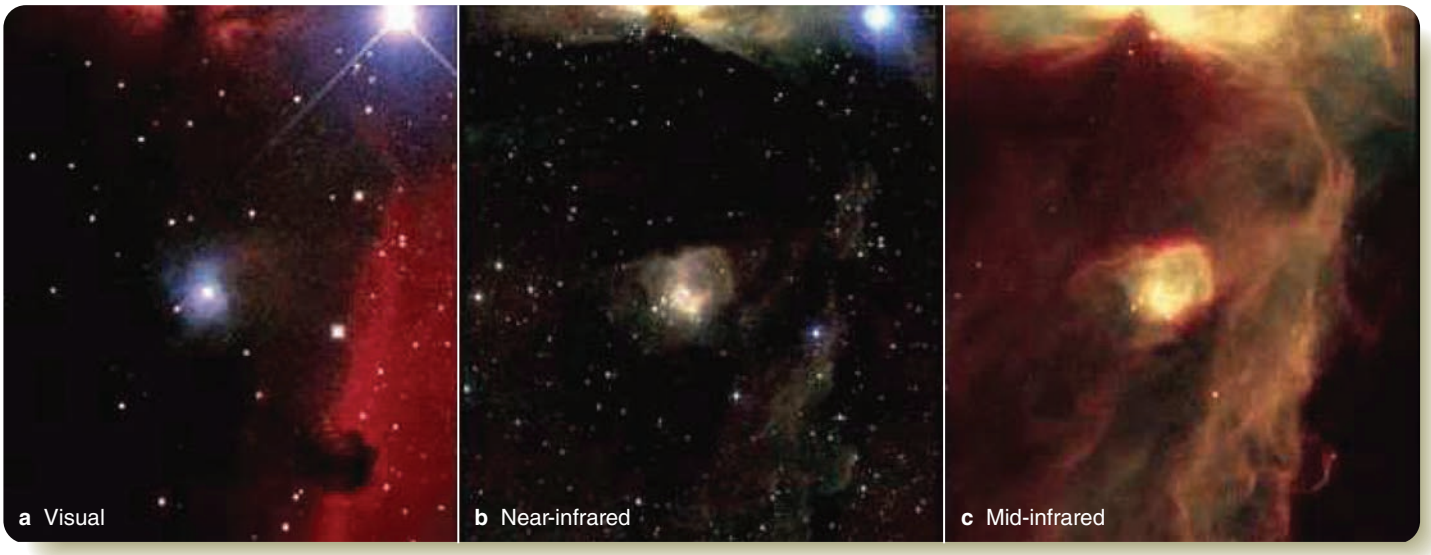
Note that interstellar reddening does not shift the wavelengths of spectral lines as does the Doppler effect. Rather, reddening changes the proportions of long- and short-wavelength photons that reach Earth, and that makes distant stars redder. You see a more ordinary example of reddening when you watch the Sun set. As the Sun approaches the horizon, sunlight must travel through a lot of air on its way to your eyes, and blue photons are more likely than red ones to be scattered by the molecules and particles in the air. A larger proportion of red photons reaches your eyes, so the setting Sun looks red.

Astronomers can measure the amount of reddening by comparing two stars of the same spectral type, one of which is dimmed more than the other. If you plot the difference in brightness between the two stars against wavelength, you get what is called a **reddening curve**. A reddening curve compares how starlight is dimmed by ISM dust at different wavelengths (**Figure 10-3**). In general, the light is dimmed in proportion to the reciprocal of the wavelength, a pattern typical of scattering from small dust particles. Laboratory measurements suggest that extra-strong extinction at wavelengths around 220 nm is caused by a spectral feature of one form of carbon, which is evidence that some of the ISM dust particles contain carbon. Near-infrared spectra of ISM dust indicate that some of the particles are hollow soccer-ball-shaped molecules containing up to 60 carbon atoms that are nicknamed "buckyballs" after inventor and architect Buckminster Fuller.

Even the less dense parts of the ISM dim starlight significantly, and some clouds of interstellar matter are so dense that they completely block our view at visual wavelengths. Look again at Figure 10-3 and notice how strong the extinction is at visible and ultraviolet wavelengths around 500 nm relative to infrared wavelengths (left side of plot). Even low-density parts of the ISM dim starlight significantly, and some clouds of interstellar matter are so dense that they completely block our view at visual wavelengths (**Figure 10-4a**). The ISM dust is quite transparent at near-infrared wavelengths (**Figure 10-4b**). At even longer mid-infrared wavelengths, the ISM glows with blackbody radiation emitted by the dust (**Figure 10-4c**).

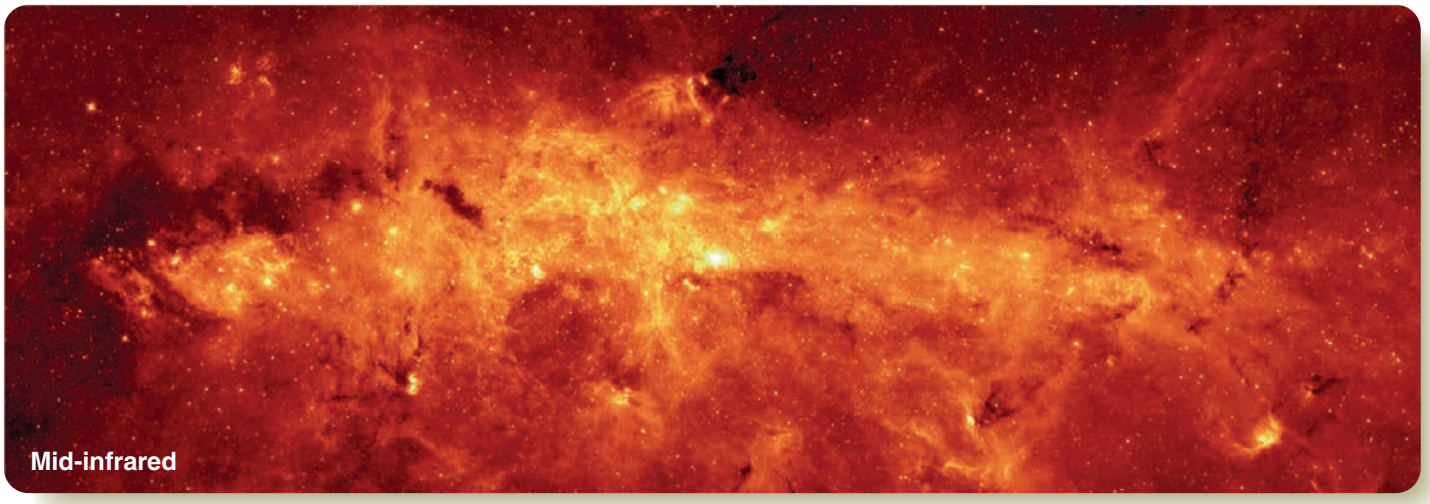
10-1c Infrared Radiation from ISM Dust

The dust in the ISM makes up only about 1 percent of its total mass, and it is very cold, with a temperature of 100 K (−280°F) or lower. Nevertheless, it is easy to detect in the infrared.



(a) Courtesy Howard McCallon; (b) 2MASS/JPL-Caltech/U. Massachusetts; (c) ESA/NASA/JAXA/ISO; 3-panel presentation assembled by Robert Hurt

▲**Figure 10-4** Images of the Horsehead Nebula at different wavelengths. (a) At visual wavelengths, the Horsehead Nebula is a dark peninsula extending away from a larger dark cloud. (b) At near-infrared wavelengths the dark nebulae are nearly transparent and stars are visible through them. (c) At longer mid-infrared wavelengths, the clouds glow with blackbody radiation emitted by dust grains.



NASA/JPL-Caltech/S. Stolovy

▲**Figure 10-5** The *Spitzer Space Telescope* imaged the center of the Milky Way Galaxy at a wavelength of 8 microns (8000 nm) to reveal the distribution of the dusty ISM. The brightest regions are clouds heated by young, massive stars. Individual stars can be seen as bright spots. Some of the clouds are so large, dense, and cold that they are opaque and dark even at this infrared wavelength.

Although the dust is cold, it is not at absolute zero, so it must radiate blackbody radiation; at those temperatures, the radiation is emitted at infrared wavelengths.

You might wonder how such cold dust can be a powerful source of infrared radiation, but you also may recall that the luminosity of an object depends on its surface area as well as its temperature (Stefan-Boltzmann law, Eq. 7-1). To illustrate, imagine a solar mass of interstellar dust. It would amount to more than a billion trillion trillion trillion (10^{45}) dust specks, each 1 micron (1000 nm, or 0.001 mm) in diameter or smaller. That much dust would have a surface area more than 10^{15} times larger than the Sun's, which means that even though it is very cold, the dust can radiate a huge amount of infrared radiation.

Infrared telescopes can reveal the ISM by mapping the energy emitted by dust. Where the gas and dust are denser or where stars warm the dust, it is more luminous. **Figure 10-5** shows part of our Milky Way Galaxy aglow with radiation from dust at an infrared wavelength of 8 microns (8000 nm).

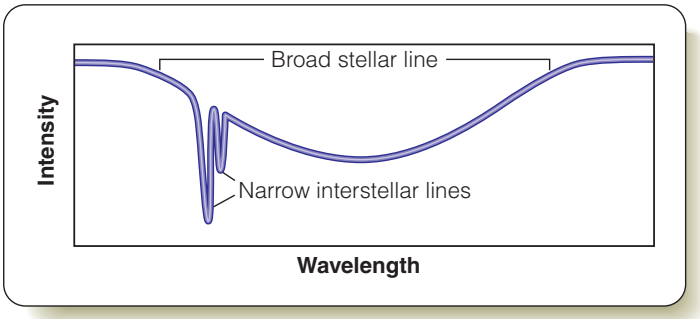
10-1d Interstellar Absorption Lines

As starlight travels through the ISM, gas atoms can absorb specific wavelengths and form **interstellar absorption lines** in the spectra of those stars. Astronomers can distinguish interstellar absorption lines from stellar absorption lines in a number of ways.

Some stellar spectra contain absorption lines that obviously don't belong to the stars because they represent the wrong ionization state. For example, if you looked at the spectrum of a very hot star such as an O star, you would expect to see no lines of once-ionized calcium because that ion cannot exist at the high temperatures found in the atmosphere of such a hot star. But many O star spectra contain spectral lines of ionized calcium. These lines must have been produced in cool material between the star and Earth rather than in the star's atmosphere (**Figure 10-6**).

The widths of the interstellar lines also give away their identities. In the atmosphere of a star, the gas is so dense that atoms collide with each other often and disturb the atomic energy levels. That blurs the spectral lines and makes them broader. Also, in a hot gas the atoms move so rapidly that they produce Doppler shifts, some red and some blue, which broaden the spectral lines. In contrast, interstellar gas is cold and has a very low density, so Doppler shifts are small and atomic collisions are rare. As a result, interstellar absorption lines are extremely narrow.

Interstellar lines are often split into two or more components. The multiple components have slightly different



▲**Figure 10-6** Interstellar absorption lines can be recognized because they are very narrow. In this plot a spectral line of singly ionized calcium produced in the atmosphere of a star is so magnified that it is seen as a broad dip in the plot. In contrast, the absorption lines produced by two clouds in the foreground ISM are much narrower.

wavelengths and must have been produced when the light from the star passed through different clouds of gas on its way to Earth. Because individual clouds of gas have slightly different radial velocities, they produce absorption lines with slightly different Doppler-shifted wavelengths.

10-1e Interstellar Emission Lines

You expect from Kirchhoff's second law (Section 7-3, especially the Concept Art: Atomic Spectra) that an excited low-density gas viewed against an empty or colder background should produce an emission spectrum, and the ISM does exactly that (for example, Figure 1 in Concept Art: Three Kinds of Nebulae). You also learned from Kirchhoff's laws that the wavelengths of a given atom or molecule's emission lines are the same as the wavelengths of its absorption lines. Emission lines, like absorption lines, give astronomers powerful ways to study the gas and dust between the stars.

Certain lines in the spectra of emission nebulae are called **forbidden lines** because they are almost never seen in the spectra of excited gas on Earth. That is because electron jumps between certain energy levels are extremely unlikely. A normal transition might occur after 10^{-8} to 10^{-7} second, but if an electron gets caught in one of these **metastable levels**, it may be stuck for as long as an hour before it falls to a lower level and emits a photon of the corresponding energy. Unless the gas has an extremely low density, collisions between atoms will disturb electrons caught in metastable levels long before they can make a transition downward and emit a photon. That means forbidden lines are produced only by very low-density gas and are generally not visible in laboratories on Earth where gas densities can't be made low enough.

But, even in relatively dense interstellar nebulae, the gas has such a low density that an atom can go for an hour or more between collisions, giving an electron stuck in a metastable level time to decay to a lower level and emit a photon at a so-called forbidden wavelength. Note that gases in Earth's upper atmosphere have low enough density to emit some forbidden lines that are visible as auroras (look back to Chapter 8).

Good examples of forbidden lines are the ones at wavelengths of 495.9 nm and 500.7 nm, which your eye sees as green, produced by oxygen atoms that have lost two electrons. (Following the convention for naming ions, twice-ionized oxygen is symbolized "O III," capital O plus Roman numeral III. Neutral oxygen is symbolized as O I.) The oxygen ions can become excited by collision with a high-energy photon or a rapidly moving ion or electron, and atoms can emit photons of various wavelengths as the electrons cascade back down to lower energy levels. Some of those electrons get caught in metastable levels and eventually emit photons at forbidden wavelengths.

The forbidden lines are clear evidence that nebulae have very low densities. This is a prime example of how astronomers can use knowledge of atomic physics gained in laboratories on Earth to understand astronomical objects; in this case, to study the ISM.

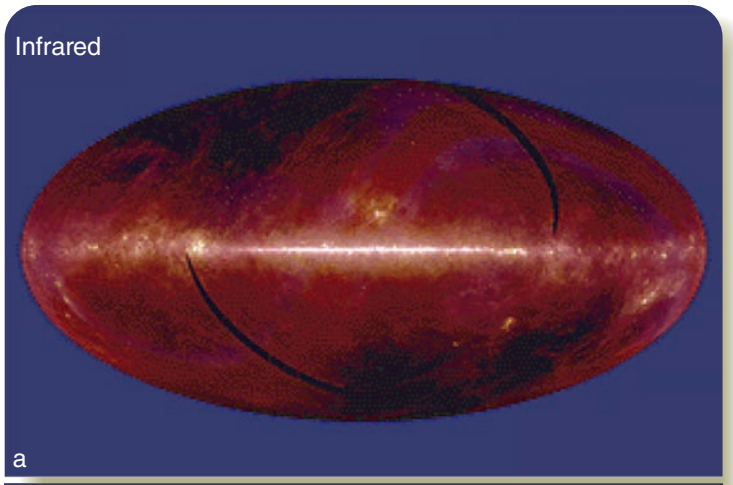
Other atoms and molecules in the ISM produce emission lines in the infrared and radio parts of the spectrum. A molecule can store energy by rotating at different speeds, by vibrating as if the atoms were linked by little springs, or by twisting back and forth. When an atom makes a transition from a high-energy state to a lower-energy state, it can emit the excess energy as a photon. The carbon monoxide (CO) molecule is an especially powerful emitter of radio energy, as is hydroxyl (OH). You learned in Section 7-3 that the hydrogen Balmer lines lie in the visible and near-ultraviolet part of the spectrum, but other series of hydrogen lines lie in the infrared. Molecular hydrogen (H_2) and water also produce strong emission lines in the infrared. Silicate (SiO) molecules contained in dust grains produce bands of strong infrared emission lines. Polycyclic aromatic hydrocarbons (PAHs) are organic molecules found in some interstellar dust grains, but also sometimes floating freely in space. Strong infrared emission from PAHs allows astronomers to study conditions in the ISM throughout our Galaxy and in other galaxies.

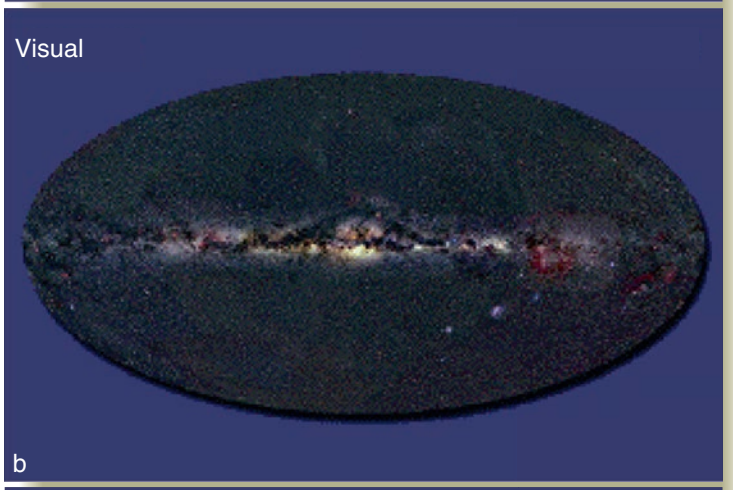
Other emission lines are found in the ultraviolet and X-ray part of the spectrum. In general, high-energy photons are produced by high-energy events, so X-ray observations tell you about very hot, excited gas such as matter expelled by the supernova explosions of stars. Infrared and radio emission lines generally tell you about colder, less excited gas such as frigid gas clouds quietly drifting through space (Figure 10-7). Observing at a wide range of wavelengths, astronomers have detected more than 150 different molecules in the ISM.

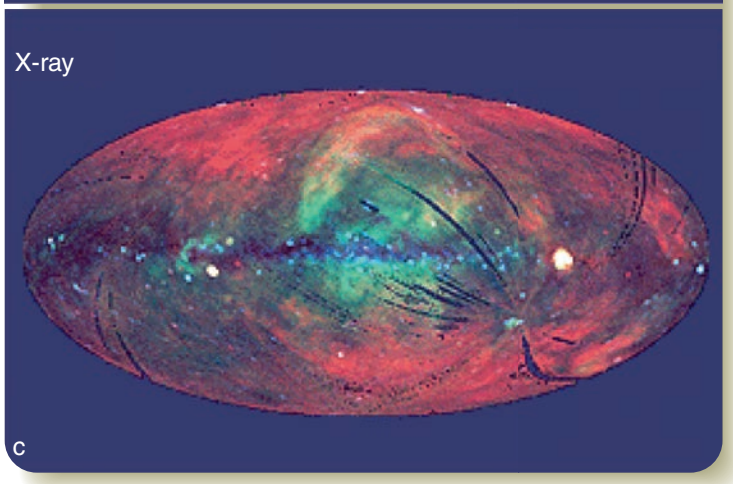
10-1f 21-Centimeter Radio Emission

When astronomers first began using radio telescopes in the 1930s, they detected a loud "hiss" at a wavelength of 21 cm coming mostly from the plane of the Milky Way. That **21-cm radiation** is an emission line from cold, neutral hydrogen gas in the interstellar medium.

How can hydrogen atoms emit radio wavelength photons? A hydrogen atom consists of an electron and a proton, and they both spin. Because they have electric charges, that spinning generates tiny magnetic fields; if the two particles spin in the same direction, the magnetic fields are reversed because the particles have opposite charges. If they spin opposite to each other, the magnetic fields are aligned. In cold, unexcited hydrogen, there are two ways the electron can settle into the ground state—spinning one way or the other way—and that means the ground state is really two energy levels separated by

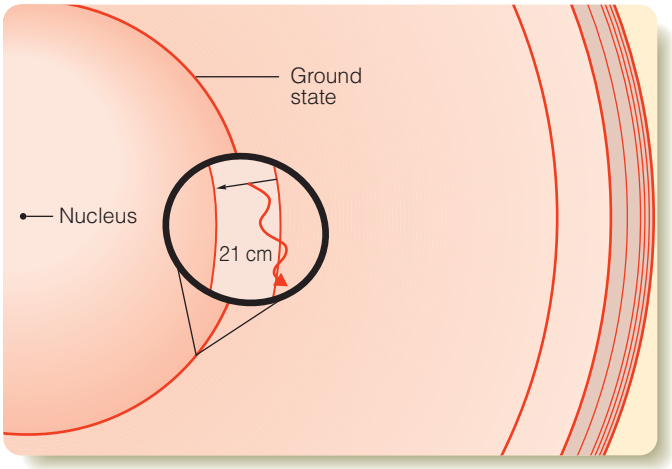






NASA

▲**Figure 10-7** These images of our Galaxy display the entire sky spread into an oval. (a) The infrared image shows the location of clouds of cold interstellar gas and dust detectable by blackbody radiation from dust grains. (b) In the visual-wavelength image, dense, opaque dusty clouds in the ISM are silhouetted against distant stars. (c) The X-ray image shows the location of very hot gas produced in most cases by the supernova explosions of dying stars.



▲**Figure 10-8** Because the proton and electron in a hydrogen atom can spin in the same or in the opposite directions, the ground state is actually two closely spaced energy levels. In this diagram, only by magnifying the image can you distinguish the two levels. When an atom decays from the upper level to the lower level, it emits a photon with a wavelength of 21 cm (not drawn to scale).

a tiny amount of energy (**Figure 10-8**). If the electron is spinning in the opposite direction to the proton's spin, it can flip over and spin in the same direction, releasing the excess energy as a photon with a wavelength of 21 cm.

The 21-cm radiation is another example of a forbidden line. That transition is statistically very unlikely, and it isn't detected in laboratories on Earth because the gas is too dense and the atoms collide too often. In space, however, hydrogen atoms collide only rarely, and hydrogen atoms can go undisturbed for millions of years, allowing them to produce the 21-cm emission line. Astronomers can use that line to map the location of the cold, low-density gas in our Galaxy (**Figure 10-9**). Also, radio waves have such long wavelengths that they are not blocked by the tiny dust grains in the ISM, allowing astronomers to observe distant clouds that would be undetectable at shorter wavelengths.

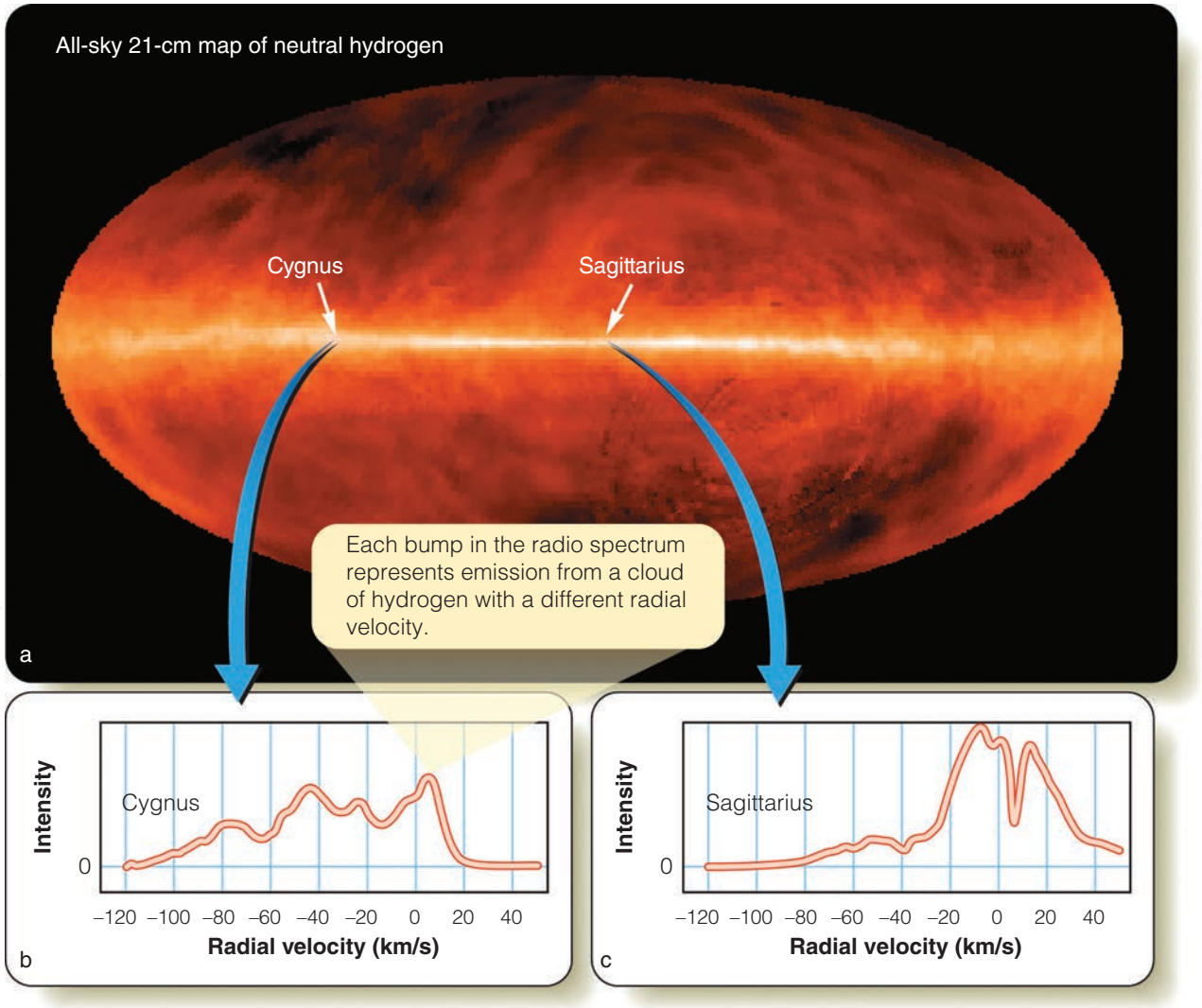
Practicing **SCIENCE**

What is the evidence that an interstellar medium exists?

Everything a scientist does is based on evidence, either acquiring observational evidence, or confronting hypotheses with evidence.

As to whether the ISM exists, there is the simple fact that parts of it are visible as nebulae. Second, detailed analysis of emission and absorption lines shows that interstellar space is filled by cold, thin gas. Third, observations of extinction and reddening of starlight show that some of the ISM must be in the form of tiny dust specks.

Try a related question: ***What do infrared observations reveal about the ISM?***



J. Dickey (U. Minnesota), F. Lockman (NRAO), SkyView

▲**Figure 10-9** (a) Most of the neutral hydrogen in the Milky Way Galaxy is in the plane of the Galaxy's disk—the bright band running from left to right across the center of this map. Notice how wispy and irregular the hydrogen clouds are. (b, c) Many clouds of neutral hydrogen orbiting the center of the Galaxy, each with its individual Doppler shift, are revealed by 21-cm radio observations.

10-2 Components of the Interstellar Medium

Astronomers can observe the ISM at many different wavelengths, and those observations are the evidence against which their hypotheses can be tested (**How Do We Know? 10-1**). The observations clearly show that the ISM is not uniform but is made up of different components.

10-2a Cool Clouds

The nebulae you see near hot, luminous stars are only small portions of the ISM. Like the parts of a fog cloud near streetlights, you see them easily because they are illuminated.

As you learned previously, the multiple interstellar absorption lines seen in stellar spectra alert you that some of the ISM consists of clouds (Figure 10-6), and observations at infrared and radio wavelengths dramatically reveal clouds of cool gas and dust with densities ranging up to about 100 atoms per cubic centimeter (Figures 10-5 and 10-9a). Because these clouds are not ionized, they are called **H I clouds**. They are typically 50 to 150 pc in diameter and have masses of a few solar masses. The gas temperature is only about 100 K. Although it is easy to think of these clouds as more or less spherical blobs of gas, observations show that they are usually twisted, flattened, and tangled into chaotic shapes, as you can see in Figure 10-1b, which is further evidence that the ISM is not static and motionless.

10-2b The Intercloud Medium

The space between the cool H I clouds is filled by a hot **intercloud medium** with a temperature of a few thousand K and

Separating Facts from Hypotheses

When scientists disagree, what do they debate?

The fundamental work of science is testing hypotheses by comparing them with facts. Facts are evidence of how nature works, and they represent reality. Hypotheses are attempts to explain how nature works. Scientists are careful to distinguish between the two.

Scientific facts are those observations or experimental results of which scientists are confident. Ornithologists might note that fewer mountain thrushes are returning each spring to a certain mountain valley. Counting bird populations reliably is difficult and requires special techniques, but if the scientists made the observations correctly, they can be confident of their result and treat it as a fact, especially if other scientists repeat the observations and get similar enough results.

To explain the declining population of thrushes, the scientists might consider a number of hypotheses, such as global warming or chemical pollution in the food chain. The ornithologists are free to combine or adjust their hypotheses to better explain the bird migration, but they are not free to adjust their facts. Scientific facts are the hard nuggets of reality that can't be changed.

New facts can aggravate debates that are already politically charged. The declining number of mountain thrushes, for example, could be unwelcome news because addressing the root problem might cost taxpayer dollars or hurt local business interests. Nonscientists sometimes debate an issue by trying to adjust or even deny the facts, but scientists are not free to ignore a fact because it is unpopular or inconvenient. Scientists debate an issue by arguing about which hypotheses best apply or how a hypothesis could be adjusted to fit the observed facts, but once established and confirmed, the facts themselves are not in question.

Whether scientists are measuring the density of an emission nebula or the size of a bird population, the final data become the reality against which hypotheses are tested. When Galileo said we should "read the book of nature," he meant we should consult reality as the final check on our understanding.

Michael A. Seeds

In science, evidence is made up of facts that can range from precise numerical measurements to a qualitative description of a flower's shape.

a density of only about 0.1 atom/cm^3. The gas is partially ionized by the ultraviolet photons in starlight, but although it is hot, it has such a low density it does not emit noticeable visible light.

Interactions between adjacent regions of the ISM are mostly determined by their relative pressures. The pressure in a gas depends on both its density and its temperature. (Note that pressure and density are related, but they are not the same.) The H I clouds are normally cool and dense, whereas the gas of the intercloud medium is hot but low in density (Table 10-1). An H I cloud and the adjacent intercloud medium can thus have similar pressures even though their temperatures are quite different. That allows the H I clouds and the intercloud medium to coexist in stable equilibrium such that neither region expands nor contracts.

Table 10-1 Four Components of the Interstellar Medium

Component	Temperature (K)	Density (particles/cm^3)	Gas State
Molecular clouds	10–50	10^3–10^5	Molecules
H I clouds	50–150	10^0–10^3	Neutral hydrogen; other atoms ionized
Intercloud medium	10^4	10^{-1}–10^0	Partially ionized
Coronal gas	10^6–10^7	10^{-4}–10^{-2}	Highly ionized

If you are thinking carefully about this, you may be wondering how the intercloud medium could be ionized when it is not close to hot stars. To understand how, imagine that you are watching an atom floating in the intercloud medium. Ultraviolet photons from distant stars are not common, but they do whiz past you now and then. Soon the atom absorbs one of these photons and becomes ionized by losing an electron. In a denser gas, it would quickly find and capture another electron, becoming a neutral atom again. However, the ISM has such a low density that the gas particles are spread far apart. Your atom must wait a very long time before an electron comes close enough to capture. That is why the gas of the intercloud medium is ionized; ions spend almost all their time waiting for an electron they can capture.

10-2c Molecular Clouds

Radio and infrared observations of the ISM reveal some clouds of gas and dust with densities ranging from hundreds to a thousand particles (atoms and molecules) per cubic centimeter. These clouds are called **molecular clouds** because they are dense enough to form molecules deep inside where the dust protects the molecules from ultraviolet photons, which would break up the fragile molecules.

Molecular clouds are very cold, only a few degrees above absolute zero, and much of the gas is in the form of molecules. Molecular hydrogen (H_2) is the main gas, but that molecule is difficult to detect, so astronomers typically study the clouds by studying radio emission from other molecules such as carbon

monoxide (CO). Although CO is 10,000 times less common in the clouds than H_2, the CO molecules are a strong source of emission at a wavelength of 2.6 mm, in the microwave part of the radio spectrum. More than 150 other molecules including hydroxyl (OH), nitrous oxide (N_2O), and ethyl alcohol (C_2H_6O) have been detected in molecular clouds, but the overall elemental composition of the clouds is the same as the ISM in general; they are about 70 percent hydrogen by mass, 28 percent helium, and 2 percent heavier atoms. The difference is that they are dense enough and cold enough for the atoms to link together and form molecules.

The CO and dust in molecular clouds tend to keep them cold. CO molecules and dust grains are very good at radiating infrared radiation, and such clouds are transparent to infrared wavelengths. So, the infrared photons zip out of molecular clouds carrying away energy, and that keeps the temperature low inside the clouds.

Molecular clouds contain up to a few hundred solar masses, but **giant molecular clouds** contain up to a million solar masses. They can be 100 pc (300 ly) in diameter and in their densest cores can contain 100,000 particles per cubic centimeter. The giant molecular clouds tend to form from gas that has settled and concentrated toward the plane of the Galaxy. You can see a number of them as dark nebulae that obscure stars and form the Great Rift along the center of the Milky Way (look again at Figure 3 in the Concept Art: Three Kinds of Nebulae). Thousands of giant molecular clouds exist in the Milky Way.

Giant molecular clouds are the nests of star birth. Deep inside these great clouds, gravity can pull the matter inward and create new stars (**Figure 10-10**). That is a story you will follow in detail in the next chapter. For now, you can complete your quick tour of the ISM by examining its hottest component.

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

▲ **Figure 10-10** The dark lower and right-hand margins of this image show part of a Giant Molecular Cloud named N90 located at a distance of about 200,000 ly in a satellite galaxy of the Milky Way. The cloud material is dense and dusty, blocking our view at visual wavelengths into and behind the cloud. In the center of the image is an H II region and reflection nebula containing a cluster of young stars that condensed from the cloud material.

10-2d Coronal Gas

The ISM would seemingly be very cold, so you might not expect X-rays to be emitted by the ISM. Nevertheless, X-ray telescopes above Earth's atmosphere have detected X-rays from parts of the ISM. This gas has been called the **coronal gas** because it has temperatures of 10^6 K or higher, similar to the Sun's corona. Of course, the coronal gas in the ISM is not related in any way to the corona of the Sun.

You might expect the coronal gas to have a very high pressure because it is so hot, but actually it has a very low density and therefore a low pressure. The coronal gas contains only 10^{-4} to 10^{-2} particle/cm^3. That is, you would have to search through a few hundred to a few thousand cubic centimeters of coronal gas—about the volume of a shoebox—to find a single particle (an ionized atom or an electron). Because of its very low density, the pressure in the coronal gas can be roughly the same as in the H I clouds, and the components of the intercloud medium can stay in equilibrium (not expand or contract).

Astronomers have evidence that most coronal gas is produced when massive stars die violently in supernova explosions (you will learn more about those in a later chapter). Although they are rare, such explosions blast large amounts of very hot gas outward in expanding bubbles. Gas flowing away from very hot, young stars can also produce coronal gas (**Figure 10-11**). In some

X-ray: NASA/CXC/CfA/R. Teullman et al.; Visual: NASA/ESA/STScI/AURA/NSF

▲ **Figure 10-11** On the right side of this combined X-ray and visual-wavelength image, as many as 200 hot, luminous, newly formed O and B stars are inflating bubbles of coronal gas that emit X-rays, represented here as blue. On the left are clouds of expanding coronal gas produced by other hot stars plus several supernova explosions. Intense star formation within this cloud is blowing it apart.

cases, neighboring bubbles of coronal gas may expand into each other and merge to form larger volumes called superbubbles. Such regions are important because the expanding coronal gas may compress nearby cool clouds and trigger star formation, as you will see in the next chapter.

In summary, observations at many wavelengths show that the ISM is complex, with four different components (look again at Table 10-1). The exact proportions are not well known, but cool clouds probably make up very roughly 20 percent of the mass and the intercloud medium about 50 percent. The molecular clouds are dense and make up about 20 percent of the mass, and the low-density coronal gas amounts to only about 10 percent of the mass in the ISM, although it accounts for more than 20 percent of the volume. In fact the four-component picture is a simplified model of the real situation, but like other useful models you have already encountered, it is adequate to portray the general properties of the ISM.

Far ultraviolet and X-ray observations confirm that the Sun is located just inside a region of coronal gas roughly 100 pc in diameter filled with hot, ionized hydrogen. This **local bubble** (or *local void*) of gas in which the Sun is located was probably inflated when a supernova exploded near the Sun within the past few million years.

The Sun "exhales" gas and dust in the solar wind (Section 8-1d). The *Voyager* probes are more than 100 AU from the Sun, heading toward interstellar space. Measurements made by the instruments on the two craft indicate they are gradually passing from the heliosphere—the region dominated by material coming from the Sun—into the ISM (**Figure 10-12a**). In addition, the *Interstellar Boundary Explorer (IBEX)* probe is able to map the heliosphere's boundary by detecting atoms sent toward Earth from collisions between the solar wind and the ISM. *IBEX* observations show that the Solar System is carving a cometlike trail through the surrounding material (**Figure 10-12b**).

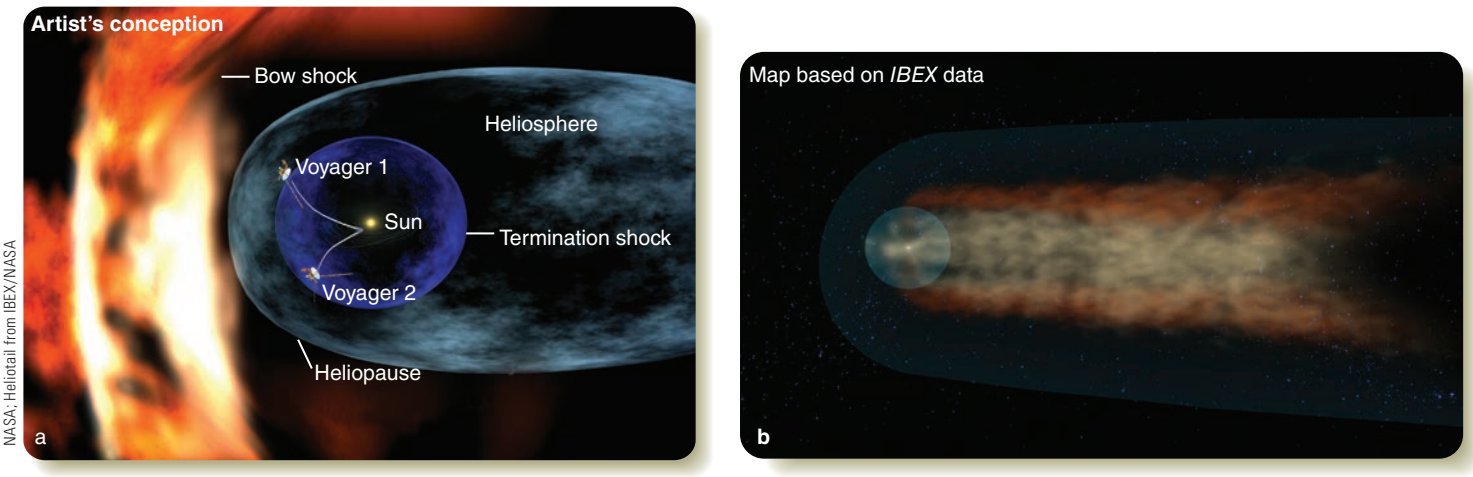
Solar System material—dust, gas, spacecraft—can end up in the ISM. Can ISM material reach Earth? Scientists analyzing comet dust captured by the *Stardust* spacecraft found at least seven grains that are unlike any known Solar System dust particles. The directions from which these particles entered the trap apparatus indicate that they probably came from the ISM. These grains evidently survived passage through the bow and termination shocks only to collide with *Stardust* and be shipped to Earth in the sample capsule.

Practicing SCIENCE

If hydrogen is the most common molecule, why do astronomers depend on the carbon monoxide (CO) molecule to map molecular clouds? This question points at the common practice of scientists inferring what they want to know from what can be more easily observed.

Although hydrogen is the most common atom in the Universe and molecular hydrogen (H_2) the most common molecule, molecules of hydrogen do not radiate strongly in the radio part of the electromagnetic spectrum. In contrast, the much less common CO molecule is a very efficient radiator of radio energy, so radio astronomers use it as a tracer of molecular clouds. Based on what is known about the chemical composition of stars and the ISM, when radio telescopes reveal a great cloud of CO, astronomers confidently infer that most of the gas is H_2. They are also confident that the molecular cloud contains dust because without the dust protecting molecules in the cloud from ultraviolet radiation the molecules would be broken into atoms.

Now use the physics of infrared emission by dust to answer a different question that lets astronomers make further inferences about the ISM: ***How can a relatively small mass of cold dust radiate vast amounts of infrared radiation?***



NASA; Heliotail from IBEX/NASA

▲**Figure 10-12** (a) Artist's conception of the two *Voyager* spacecraft trajectories in relation to the Solar System and the ISM. The central heliosphere that is completely filled by the solar wind extends to the heliopause. The wind slows down as it begins to interact with the ISM at the boundary labeled *termination shock*. The bow shock is the outermost boundary between the Solar System and the ISM; it is curved much like the bow wave in front of a speedboat because of the Solar System's motion toward the left of the diagram. *Voyager 1* apparently crossed the bow shock in mid-2012, and *Voyager 2* is expected to cross it sometime before 2020. (b) Portrait of the heliotail, the trail made by the heliosphere as it moves through the ISM, based on measurements made by the *Interstellar Boundary Explorer (IBEX)* "atomic telescope" in orbit around Earth.

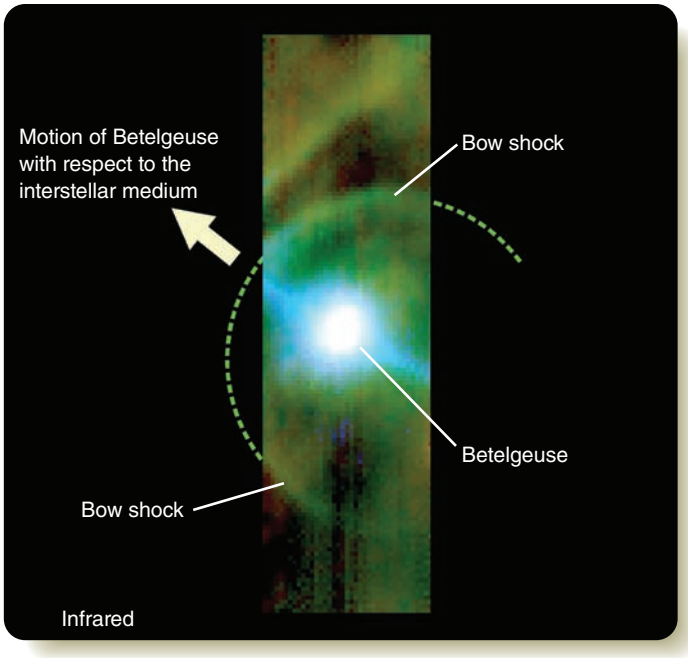
10-3 The Gas-Stars-Gas Cycle

The ISM is complex, and in some cases, quite beautiful, but it would be just a curiosity if it were not intimately related to stars. Stars are born from the ISM, and aging or dying stars return gas and dust back into the ISM. As you learn about that cycle, you are beginning to learn about the life story of the stars that will be portrayed in the next few chapters.

10-3a Gas and Dust from Aging Stars

Astronomers have clear evidence that much of the hydrogen and helium in the ISM is gas left over from the formation of our Galaxy more than 10 billion years ago, so that gas has never been inside stars. Some of the gas and dust that lie between the stars, however, has been produced in stars and expelled as the stars aged.

The Sun's wind is a gentle breeze compared to the wind blown out of some aging stars. The atmospheres of those stars are so cool that certain types of atoms, including carbon, silicon, and oxygen, can condense to form specks of dust. The pressure of the starlight pushes those dust specks out of the star along with a powerful flow of hydrogen, helium, and other gases. Infrared observations of Famous Star Betelgeuse reveal that, as it moves through space, its powerful wind plows a trail through the ISM similar in shape to the Sun's (Figure 10-12) but with a bow shock that is 1000 times larger (**Figure 10-13**). The gas and dust in stellar winds is destined to be added to the ISM.



From Ueta et al. 2008, PASJ

▲**Figure 10-13** The red supergiant star Betelgeuse is expelling a powerful wind of gas and dust. As is the case for the Sun, the wind from Betelgeuse forms a bow shock where it collides with the ISM that allows astronomers to infer properties of both the star's wind and the ISM. The material in stellar winds ultimately becomes new ISM.

Calculations of the Sun's future evolution indicate that it will blow away almost half of its mass in several relatively short episodes about 6 to 7 billion years from now before it is finally extinguished. In similar ways, some of the gas and dust in the ISM have been produced by previous generations of stars.

You read in the previous section that some stars die in violent supernova explosions that blast matter outward in expanding clouds of hot coronal gas. As those gases cool, dust condenses like soot from a candle flame, so supernovae also add large amounts of dust to the ISM.

Supernova explosions plus the blasts of light and gas coming from the hottest stars keep the ISM stirred with **shock waves**, the astronomical equivalent of sonic booms. As those shock waves sweep through the ISM, they compress gas and produce clouds dense enough for molecules to begin forming.

10-3b A Preview of Star Formation

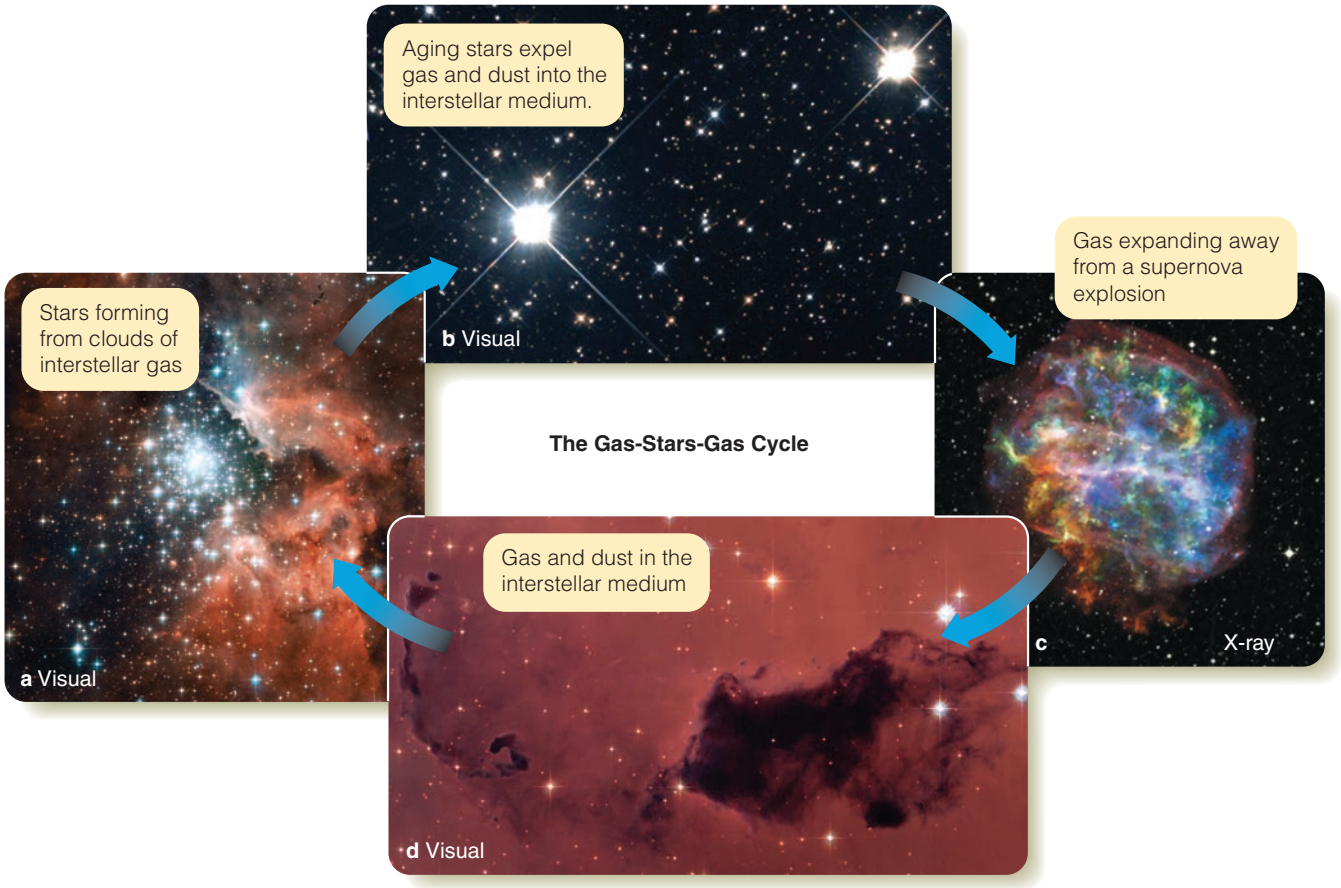
Molecular clouds form as the ISM is compressed, and gravity and turbulence tend to concentrate these clouds to form giant molecular clouds. Such clouds are located along the spiral arms of our Galaxy.

Because these clouds of gas are so cool, gravity can squeeze them to higher densities, especially if they are compressed further by passing shock waves. The densest regions of molecular clouds can become gravitationally unstable, meaning the gravity pulling the gas together is stronger than the pressure and turbulence in the gas. As the gas falls together, the densest regions break into fragments and form clusters of stars. A single giant molecular cloud can sometimes form several clusters containing hundreds or thousands of stars each.

The most massive stars to form are the hottest and expel gas most violently. Also, as you will learn in a later chapter, the most massive stars live very short lives—only a few million years—and die in violent supernova explosions blasting more gas back into the ISM.

A giant molecular cloud can make lots of stars, but as supernovae begin expelling coronal gas, the molecular cloud can be ripped apart and mixed back into the general ISM, and that closes the circle on the gas-stars-gas cycle (**Figure 10-14**).

As matter in our Galaxy passes through this cycle, some of the mass gets incorporated into low-mass stars that live very long times or gets trapped in remnants of stars like white dwarfs, neutron stars, and black holes that you will learn about in later chapters. Consequently, the matter in those remnants is removed from the cycle. Also, with each cycle some of the hydrogen gets fused into helium and heavier atoms like carbon, nitrogen, and oxygen. Thus, as the eons pass, our Galaxy is gradually using up its star-making mass and hydrogen fuel as it forms stars. Also, with each generation of stars, the ISM becomes richer in those heavier atoms, and the chemical composition of the ISM slowly changes. In a later chapter, that fact will give you a critical clue to understanding the life story of our Galaxy.


Aging stars expel gas and dust into the interstellar medium.
b Visual
Gas expanding away from a supernova explosion
Stars forming from clouds of interstellar gas
The Gas-Stars-Gas Cycle
Gas and dust in the interstellar medium
a Visual
c X-ray
d Visual

(a) ESO; (b) NASA/ESA/STScI/AURA/NSF; (c) NASA/CXC/Penn State/S. Park et al.; (d) NOAO/AURA/NSF

▲**Figure 10-14** Matter cycles from the ISM to form stars and back into the ISM. Our Galaxy is slowly but steadily using up its star-making supplies and exhausting its reserve of the hydrogen fuel that powers newborn stars.

WHAT ARE WE?

Appreciative

City dwellers see only the brightest stars scattered across the sky, but even those who live far from city lights see almost nothing between the stars. It is easy to imagine that space is empty. Only a few hazy nebulae such as the one in Orion's sword hint at the hidden richness of space.

We live in a Universe filled with beauty beyond our senses. Special telescopes and cameras can capture images that let us appreciate glowing gas clouds containing brilliant stars and dark dust lanes twisting through space. Clouds of neutral gas, cold dust, and hot bubbles of coronal gas are not visible to our eyes, but they are part of the complex beauty of the ISM. All around us is an invisible ocean—the peaceful reaches of which are blasted here and there by storms of light and heat as the ISM interacts with stars.

Science enriches our lives by giving us the means to appreciate the beauty we cannot sense directly.

Study and Review

SUMMARY

10-1 Studying the Interstellar Medium

- The **interstellar medium (ISM)**, the gas and dust between the stars, is mostly concentrated near the plane of our Milky Way Galaxy and has an average density of about one atom per cubic centimeter.
- A **nebula** is a cloud of gas and dust in space. An **H II region**, also known as an **emission nebula**, is produced when ultraviolet radiation from hot stars ionizes nearby gas, making the gas glow. The red, blue, and violet Balmer lines blend together to produce the characteristic pink-red color of ionized hydrogen.
- A **reflection nebula** is produced by gas and dust illuminated by a star that is not hot enough to ionize the gas. The dust in the cloud scatters the starlight to produce a reflection of the stellar spectrum.

Because shorter-wavelength photons scatter more easily than longer-wavelength photons, reflection nebulae look blue. The sky on a clear day looks blue for the same reason.

- A **dark nebula** is a cloud of gas and dust that is noticeable because it blocks the light from distant stars. The irregular shapes of these dark nebulae reveal the turbulence in the ISM, which shapes the cloud.
- **Interstellar dust** makes up roughly 1 percent of the mass of the ISM. The remaining 99 percent of the mass is gas. About 70 percent of the mass of the ISM is hydrogen gas, and 28 percent is helium. About 2 percent is atoms heavier than helium. This is approximately the same composition as the Sun and other stars.
- **Interstellar extinction**, or dimming, makes distant stars look fainter than they should. **Interstellar reddening** makes distant stars appear too red because dust particles in the ISM scatter blue light more easily than red light. A plot called a **reddening curve**, which shows the dependence of extinction on wavelength, indicates that the scattering dust particles are very small. The dust is made of carbon, silicates, iron, ice, and other organic compounds.
- Interstellar dust, although very cold, emits blackbody radiation. Because the dust in an interstellar cloud has a huge surface area, it can emit large amounts of long-wavelength infrared radiation.
- The interstellar gas is cold and has a very low density, and this makes **interstellar absorption lines** much narrower than the spectral lines produced in stars. Such lines are usually obvious in stellar spectra because they represent ions that cannot exist in the atmospheres of the stars. Interstellar lines stand out in the spectra of spectroscopic binaries because they do not shift their wavelengths as do lines produced by the orbiting stars. Multiple interstellar lines reveal that the starlight has passed through more than one interstellar cloud on the light's way to Earth.
- The gas in ISM nebulae has a low density, and the atoms collide so rarely that an electron caught in a **metastable level** can remain there long enough to finally fall to a lower level and emit a photon. This produces so-called **forbidden lines** that are not seen in laboratory spectra on Earth, where the atoms in the gas collide too often.
- The low-density gas of the ISM produces emission lines at many wavelengths. The **21-cm radiation** is a forbidden line produced when the electrons in hydrogen atoms change the direction of their spin, emitting radio-wavelength photons during the spin flips. This radiation allows radio astronomers to map the distribution of neutral hydrogen gas in the ISM.

10-2 Components of the Interstellar Medium

- The ISM is made up of four main components. **H I clouds** of neutral hydrogen are cold clouds containing a few solar masses in a region 10 to 150 pc in diameter. They are separated from each other by a hotter, but lower-density gas called the **intercloud medium**. The H I clouds and the medium generally have similar pressures and are thus in equilibrium.
- Large, dense, very cold clouds of gas and dust are called **molecular clouds** because they are so dense and cold that molecules can form inside them. Molecular hydrogen (H_2) is the primary constituent in these clouds but does not radiate at radio wavelengths. Hence, such clouds can be mapped using radio photons emitted by the carbon monoxide (CO) and hydroxyl (OH) molecules in the clouds. The largest of these molecular clouds are called **giant molecular clouds**, which are often the sites of star formation.
- X-ray observations have detected very hot **coronal gas**, the fourth component of the ISM. Coronal gas is hot gas blown away from massive stars and gas expelled from supernova explosions. Far-ultraviolet observations show that the Sun is located in a **local bubble** of coronal gas that was apparently produced by a supernova explosion near the Sun within the past few million years.

10-3 The Gas-Stars-Gas Cycle

- The ISM interacts with stars through a gas-stars-gas cycle. Hot stars and supernovae expel hot gas and dust into the ISM. Cool stars also contribute gas and dust. The gas and dust collect and cool, forming clouds of which giant molecular clouds are the coldest and densest, sites where new stars form singly or in clusters. Star formation ends in a molecular cloud when hot gas from newborn massive stars and from supernova explosions forms **shock waves** that blow the cloud apart.
- Not all matter in our Galaxy passes through the gas-stars-gas cycle. Some remains in low-mass stars, and some gets trapped in stellar remnants such as white dwarfs, neutron stars, and black holes. With each cycle, more and more hydrogen is converted into helium and heavier elements. As such, our Galaxy is gradually using up its star-making supplies.

REVIEW QUESTIONS

1. What evidence visible to human eyes can you cite that the spaces between the stars are not totally empty?
2. What evidence can you cite that the ISM contains both gas and dust?
3. Name the three kinds of nebula. How are they similar, and how are they different?
4. What is the difference between H I and H II?
5. If gas clouds can be in atomic, ionic, or molecular hydrogen phase, rank these cloud phases in order of coldest to warmest.
6. How do the spectra of H II regions differ from the spectra of reflection nebulae? Why?
7. How is the blue color of a reflection nebula related to the blue color of the daytime sky?
8. How much longer does an electron stay in a metastable state, compared with an ordinary energy level? Explain why an emission line produced by a transition from that metastable state is considered forbidden because of this ratio.
9. Why do some distant stars look redder than their spectral types suggest?
10. Which lines are narrower—stellar absorption spectral lines or ISM absorption spectral lines? Why?
11. Why do some spectral lines, which are forbidden in spectra on Earth, appear in spectra of interstellar clouds and nebulae? What does that tell you?
12. Why can the 21-cm radio emission line of neutral hydrogen be observed in the ISM but not easily in the laboratory?
13. If you view along a line-of-sight about 90 degrees away from the central plane of the Milky Way and look between the stars, will you likely be seeing abundant molecular gas? Atomic gas? Interstellar dust? Explain your answer.
14. Why is the ISM transparent at near-infrared and radio but opaque in visual wavelengths?
15. Name the four components of the ISM in order of hottest to coldest.
16. Of the four components of the ISM, which is the most abundant by mass?
17. How can the H I clouds and the intercloud medium have similar pressures when their temperatures are so different?
18. Why can molecular clouds form molecules in their interiors, whereas H I clouds cannot?
19. Describe a hypothesis for what produces the coronal gas. What is the evidence?

20. How does the interstellar medium interact with stars?
21. Name two sources that contribute material to the ISM.
22. Name two processes (or objects) that remove material from the ISM.
23. Does the Sun lie in a molecular H cloud, an atomic H cloud, or an ionized H cloud?
24. How do today's molecular clouds and stars differ from previous generations of stars and molecular clouds?
25. **How Do We Know?** Why are scientists free to adjust their hypotheses but not their facts?

ACTIVE INQUIRY QUESTIONS

1. When you see distant streetlights through smog generated by soot particles, they look dimmer and redder than they do normally. But when you see the same streetlights through fog or falling snow, they look dimmer but not redder. Use your knowledge of the ISM to discuss the relative sizes of the particles in smog, fog, and snowstorms compared to the wavelength of visible light.
2. If you could see a few stars through a dark nebula, how would you expect their spectra and colors to differ from similar stars just in front of the dark nebula?

PROBLEMS

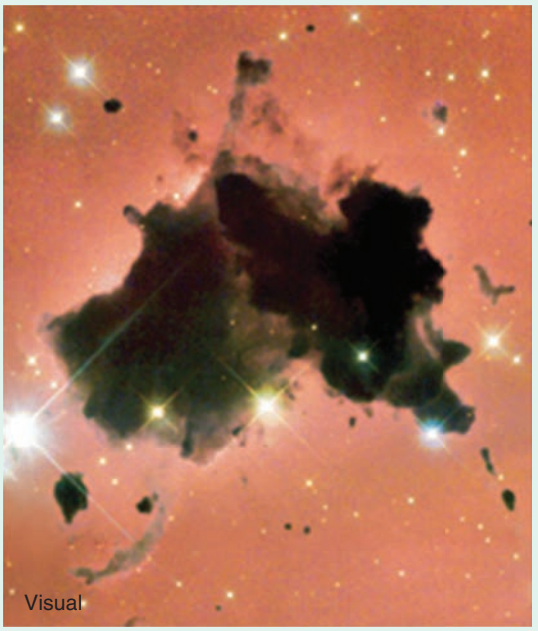
1. A small nebula has an angular diameter of 20 arc seconds and a distance of 1000 pc from Earth. What is the linear diameter of the nebula in parsecs? In meters? (*Hint*: Use the small-angle formula, Eq. 3-1.)
2. The dust in a molecular cloud has a temperature of about 50 K. At what wavelength does it emit the maximum intensity? (*Hint*: Use Wien's law, Eq. 7-2.)
3. Extinction dims starlight by about 1 magnitude per 1000 pc. What fraction of photons survives a trip of 1000 pc? (*Hint*: Consider the definition of the magnitude scale in Chapter 2. Assume the star lies in the plane of the Milky Way Galaxy.)
4. If the total extinction through a dark nebula along the telescope's line of sight is 10 magnitudes, what fraction of photons makes it through the cloud? (*Hint*: See Problem 3.)
5. The number density of air in a child's balloon is roughly the same as sea level air, 10^{19} particle/cm^3. If the balloon is now 20 cm in diameter, to what diameter would it need to expand to make the gas inside have the same number density as the ISM, about 1 particle/cm^3? (*Note*: The volume of a sphere is $\frac{4}{3}\pi r^3$.)
6. Calculate the frequency in megahertz (MHz) of the neutral hydrogen forbidden line that has an exact wavelength of 21.12 cm. Is that in the very high frequency (VHF) band from 30 to 300 MHz that includes FM radio, or in the ultra-high frequency (UHF) band from 300 to 3000 MHz that includes most TV broadcasts? (*Hint:* Refer to the relationship among wave speed, frequency, and wavelength, Chapter 6). (*Notes:* The speed of light is 3.00×10^{10} cm/s; 1 MHz $= 1 \times 10^6$ Hz.)
7. If a giant molecular cloud has a diameter of 30 pc and drifts relative to nearby objects at 20 km/s, how long does it take to travel a distance equal to its own diameter? (*Notes:* 1 yr $= 3.2 \times 10^7$ s; 1 pc $= 3.1 \times 10^{16}$ m. Remember to convert km to m.)
8. An H I cloud is 4 pc in diameter and has a density of 100 H atoms/cm^3. What is its mass in units of solar masses? (*Notes:* The volume of a sphere $= \frac{4}{3}\pi r^3$; 1 pc $= 3.1 \times 10^{16}$ m; the mass of an H atom is 1.7×10^{-27} kg; one solar mass is 2.0×10^{30} kg. Remember to convert cubic cm into cubic m.)
9. A giant molecular cloud is 30 pc in diameter and has a density of 3000 hydrogen molecules/cm^3. What is its mass in units of solar masses? (*Notes:* Useful quantities can be found in Problem 8. A hydrogen molecule consists of 2 H atoms.)
10. At what range of wavelengths does a blackbody at the temperature range of the coronal gas radiate most intensely? (*Hint:* Use Wien's law, Eq. 7-2; $\boldsymbol{\lambda max} = 2.90 \times 10^6$ nm/T (K))

SENSE OF PROPORTION

1. The Great Nebula in Orion (also known as Messier 42) is about 26 light-years across. The diameter of Neptune's orbit is about 8 light-hours. If you made model of the Solar System in which Neptune's orbit was 1 meter (40 in.) in diameter, how large on the same scale would a model of M42 be?
2. Make another model with the same scale as the one in Question 1, with 1 meter representing 8 light-hours, the diameter of Neptune's orbit. In this model, how far away is the Orion Nebula, which in reality is 1500 light-years away?

LEARNING TO LOOK

1. Examine the Lagoon Nebula on the right-hand page of Three Kinds of Nebula. Is the white object or the pink object the primary energy source? In what wavelength band of the electromagnetic spectrum is that energy emitted? How do you know?
2. Examine NGC 1973, 1975, and 1977 on the right-hand page of Three Kinds of Nebula. Are the white objects or the blue object the primary energy source? In what wavelength band of the electromagnetic spectrum is that energy emitted? How do you know?
3. A blue nebula surrounds the bright blue star in the center of Figure 10-4a. What can you conclude about the temperature of this star?
4. Look at Figure 10-6. Which atmospheric layer of the star caused the broad stellar line? Where is the light originally coming from that is being removed by the gas that caused the broad stellar line? Where is the gas that generated the narrow interstellar lines?
5. This image shows two nebulae, one pink in the background and one black in the foreground. What kind of nebulae are these?



NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

11 The Formation and Structure of Stars

NASA/ESA/STScI/AURA/NSF/A. Nota

▲ Here, a few million years ago a giant molecular cloud collapsed gravitationally, forming several massive star clusters visible near the center of the image. Intense radiation from the hottest, most luminous of the newborn stars is exciting some of the nebula's gases to create an emission nebula (pink-purple color). Notice remaining dark molecular cloud material at right and lower right.

The previous chapter introduced you to the complex and dynamic clouds of gas and dust that fill the spaces between the stars. In this chapter you will put together observations and hypotheses to learn how the interstellar medium condenses into stars, and what the conditions inside stars must be like. On the way to that understanding, you will find answers to four important questions:

- How do stars form?
- What is the evidence that stars are forming now?
- How do stars maintain their stability?
- How do stars make energy?

When you finish this chapter, you will have gained a view of the energetic childhood of stars. In the next chapter, you will follow stars through their stable, slowly evolving adult life stage called the *main sequence* and beyond that to their transformation, at the ends of their lives, into giant stars.

> Jim he allowed [the stars] was made, but I allowed they happened. Jim said the moon could'a laid them; well, that looked kind of reasonable, so I didn't say nothing against it, because I've seen a frog lay most as many, so of course it could be done.
>
> —MARK TWAIN, *THE ADVENTURES OF HUCKLEBERRY FINN*

THE STARS ARE NOT ETERNAL. When you look at the sky, you see hundreds of points of light, and, amazingly enough, each one, like the Sun, is a tremendous nuclear fusion reactor held together by its own gravity. The stars you see tonight are the same stars your parents, grandparents, and great-grandparents saw. Stars don't change perceptibly in a person's lifetime, or even during all of human history, but they do not last forever. Stars are "born," and stars "die." This chapter begins that story.

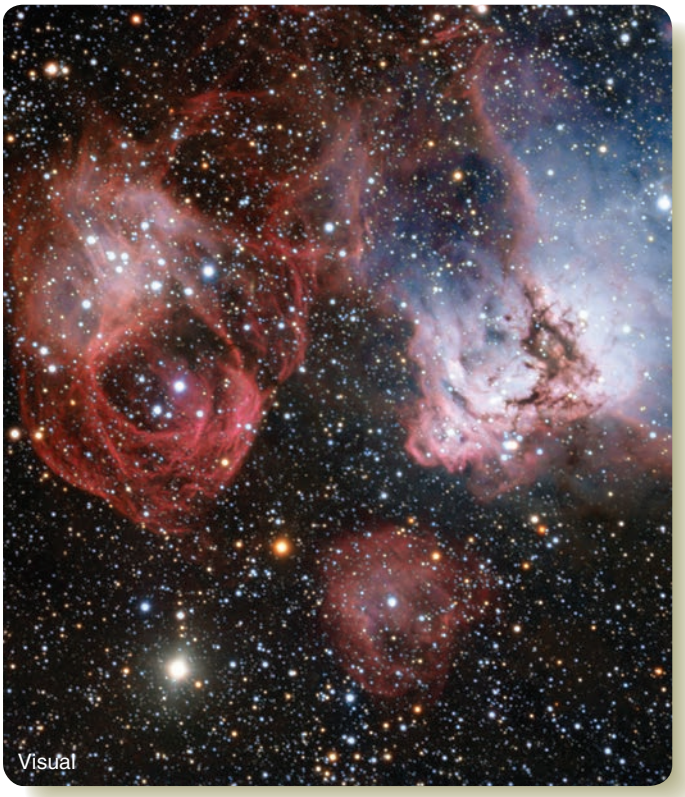
How can you know what the life cycles and internal processes of stars are, given that you won't live long enough to see them evolve, and you can't see inside them? The answer lies in the methods of science. By constructing hypotheses about how nature works and then testing those hypotheses against evidence from observations, you can unravel some of nature's greatest secrets. In this chapter, you will see how gravity creates stars from the thin material of interstellar space. You will then learn how the flow of energy from the insides of stars toward their surfaces balances gravity and makes the stars stable, and how nuclear reactions in the cores of stars ultimately supply that flowing energy. To understand this story, in your imagination you will plunge from the cold gas of the interstellar medium into the hot centers of the stars themselves.

11-1 Making Stars from the Interstellar Medium

Astronomers find hundreds of places in the sky where clouds of gas glow brightly because they are illuminated by the hottest, most massive, and most luminous main-sequence stars, with spectral types O and B. As you will learn in detail in the next chapter, those kinds of stars pour out such floods of energy that they can last only a few million years, so you must be seeing them pretty near where they were born. O and B stars are therefore signposts pointing to star-forming regions, which are always among the densest parts of the interstellar medium (ISM; **Figure 11-1**).

11-1a Star Birth in Giant Molecular Clouds

As you saw in the previous chapter, spectra of the ISM show that its chemical composition is basically the same as that of stars. By observing at far-infrared and radio wavelengths, to which interstellar dust is basically transparent, astronomers can peer inside even dense molecular clouds and find evidence that stars are forming there. In fact, the giant molecular clouds discussed in the preceding chapter are each massive enough to make thousands of new stars. But how do those large, low-density clouds of cold gas and dust become comparatively small, high-density, hot stars?



▲ **Figure 11-1** On the right side of this image is NGC 2035, a star-forming region that includes emission nebulae (*pink, center right*) and reflection nebulae (*blue, far right*) illuminated by newly formed O and B stars, plus dense, dark nebulae where stars continue to form. On the left is a supernova remnant, clouds of hot gas expanding away from an explosion marking the death of a massive star. This field of view is several hundred light-years across.

Giant molecular clouds have typical diameters of 50 parsecs (pc) and typical masses exceeding 10^5 solar masses, vastly larger than a single star. The gas in a giant molecular cloud is about 10^{20} times less dense than the gas in the center of a star and has temperatures of only a few kelvins—not millions of kelvins. Those clouds can form into stars only if small portions of the cloud can be compressed somehow to high density and high temperature. However, at least four factors cause an interstellar gas cloud to resist compression, and those factors must be overcome by gravity before star formation can begin.

First, thermal energy in the gas appears as random motions of the atoms and molecules. Even at the very low temperature of 10 K, the average hydrogen molecule moves at 0.35 km/s (about 780 mph). For a cloud to contract, its gravity must be strong enough to overcome that much thermal motion tending to make the cloud expand.

Second, the ISM is filled with a magnetic field. The ISM field is only about 0.0001 times as strong as Earth's magnetic field, but that can be enough to act as an internal spring and prevent the gas from contracting. Neutral atoms and molecules are unaffected by a magnetic field, but ions, having an electric charge, cannot move freely through a magnetic field. Although the gas in a molecular cloud is mostly neutral, some ions are included, and that means a magnetic field can exert a force on the gas as a whole. Under some conditions, the ionized gas in giant molecular clouds can gradually recombine with free electrons. The resulting less-ionized gas is freer to "slip past" the magnetic field and contract; this process has been observed inside isolated molecular clouds. Even so, the cloud's gravity must overcome resistance caused by the interstellar medium to make the bulk of the gas contract.

Third, everything in the Universe rotates. As a gas cloud begins to contract, it spins more and more rapidly as angular momentum is conserved—just as ice-skaters spin faster as they pull in their arms (look back to Figure 5-6). This rotation can become so rapid that it counteracts gravity and the cloud resists further contraction.

Fourth, the ISM is turbulent. In the previous chapter, you saw that nebulae are often twisted and distorted by strong currents. This turbulence is another effect opposing gravity that can help prevent a molecular cloud from contracting.

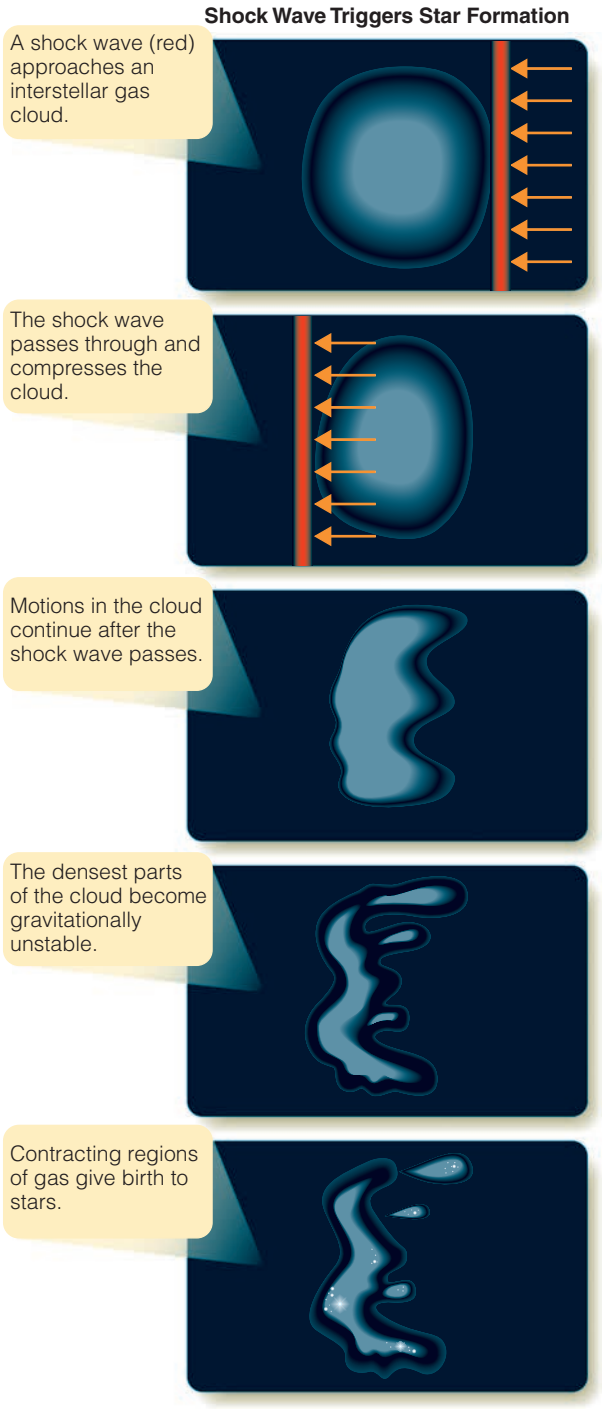
Given these four resisting factors, it may seem surprising that any giant molecular clouds can contract at all, but radio and infrared observations show that at least some giant molecular clouds develop concentrated regions called **dense cores** that are roughly 0.1 pc in size and contain about 1 solar mass. A single giant molecular cloud may contain hundreds of these dense cores, each possibly destined to become a star.

Both theory and observation suggest that giant molecular clouds can be triggered to form stars by a passing shock wave—sometimes just called a *shock* (Figure 11-2). During such a triggering event, some regions of the large cloud can be compressed to such high densities that the resisting effects can no longer oppose gravity, and star formation begins.

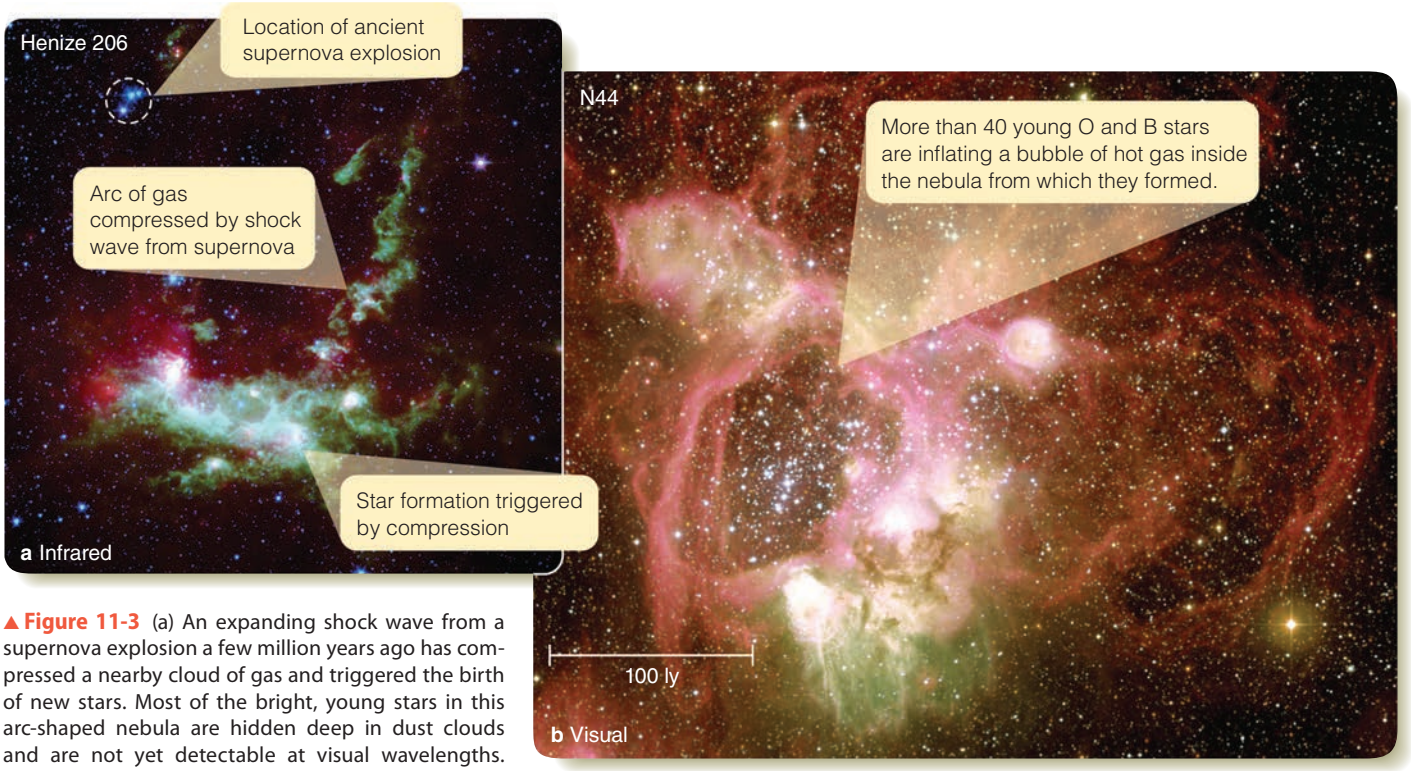
Shock waves that can trigger star formation are common in the ISM. For example, gas flows of several sorts produce shocks when they push into colder, denser interstellar matter. Supernova explosions create powerful shock waves that rush through the ISM (Figure 11-3a). Also, the ignition of very hot stars ionizes nearby gas and causes it to flow rapidly away (Figure 11-3b). New stars of all types seem to emit strong winds and jets while they are forming that can also produce shocks.

A second type of star formation trigger is the collision of molecular clouds. Because the clouds are large, they are likely to run into each other occasionally, and because they contain magnetic fields they cannot easily pass through each other. A collision between such clouds can compress parts of the clouds and cause star formation.

A third type of trigger is the spiral pattern of our Milky Way Galaxy (look back to Figure 1-11). You will learn in a later chapter that evidence and model calculations both indicate that spiral arms are shock waves that travel around the disk of a galaxy like the moving hands of a clock. As a cloud passes through a spiral arm, the cloud can be compressed, and star formation may begin. Astronomers have found regions of star formation in which each of these three trigger processes can be identified.


Shock Wave Triggers Star Formation
A shock wave (red) approaches an interstellar gas cloud.
The shock wave passes through and compresses the cloud.
Motions in the cloud continue after the shock wave passes.
The densest parts of the cloud become gravitationally unstable.
Contracting regions of gas give birth to stars.

▲ **Figure 11-2** In this summary of a computer model, an interstellar gas cloud is triggered into star formation by a passing shock wave. The events summarized here might span about 6 million years.


Henize 206
Location of ancient supernova explosion
Arc of gas compressed by shock wave from supernova
Star formation triggered by compression
a Infrared
N44
More than 40 young O and B stars are inflating a bubble of hot gas inside the nebula from which they formed.
100 ly
b Visual

Panel a: NOAO/AURA/NSF/ Panel b: V. Gorjian; ESO

▲ **Figure 11-3** (a) An expanding shock wave from a supernova explosion a few million years ago has compressed a nearby cloud of gas and triggered the birth of new stars. Most of the bright, young stars in this arc-shaped nebula are hidden deep in dust clouds and are not yet detectable at visual wavelengths. (b) Nebula N44 has given birth to a large cluster of stars. The process of star formation can spread to new regions by the effects of stars already formed.

A single giant molecular cloud containing more than a hundred thousand solar masses does not contract to form a single enormous star. Evidently, collapsing clouds break into fragments, and the densest parts form a number of dense cores. Exactly why and how clouds divide into fragments isn't fully understood, but their rotation, magnetic fields, and turbulence apparently play important roles. Whatever the reason, when a giant cloud of interstellar matter contracts, it forms many new stars simultaneously.

11-1b Heating by Contraction

You can understand how low-density clouds of interstellar gas might contract and become dense enough to make stars, but how does the cold gas become hot enough to be a star? The answer, once again, is gravity.

To see how gravity can heat the gas, shift your attention to one dense cloud core destined to become a single star. Once that small cloud of gas begins to contract, gravity draws the atoms toward the center. This means the atoms are falling, and they gain speed as they fall. In fact, astronomers refer to this early stage in the formation of a star as **free-fall collapse**. Although the atoms may have been moving slowly to start with, by the time they have fallen most of the way to the center of the cloud, they are traveling rapidly.

Thermal energy is the agitation of the particles in a gas, so this increase in speed as a result of free-fall collapse is a step toward heating the gas. However, you can't say that the gas is hot simply because all of the atoms are moving fast. The air in the cabin of a jet airplane is traveling rapidly, but it isn't hot because all of the atoms are moving in generally the same direction along with the plane. The high speed of the infalling atoms is not converted into thermal energy until their motion becomes randomized, and that happens when the atoms begin to collide with one another as they fall into the central region of the cloud. The jumbled, random, rapid motion resulting from those collisions is thermal energy, and so the temperature of the gas increases.

This is an important principle in astronomy. Whenever a cloud of gas contracts, the atoms move "downward" in the gravitational field, pick up speed, and collide more rapidly, and so the gas grows hotter. Astronomers express this by saying that gravitational energy is converted into thermal energy. When a gas cloud expands, gas atoms move "upward" against gravity and lose speed, and the gas becomes cooler. Astronomers say that in this case thermal energy is converted into gravitational energy. The principle applies not only to clouds of interstellar gas but also to contracting and expanding stars, as you will see in the following chapters.

Previously you learned that thermal energy in the form of random atomic motion is one of the factors that can resist gravity and keep a cloud from contracting. So, you may ask, does the contraction stop once the free-fall collapse heats the gas? The answer is no, not if there is a way for the compressed cloud to get rid of thermal energy by radiating it away. The contracting cloud is still relatively cold; according to Wien's law (Eq. 7-2), its material should produce very long-wavelength radiation to which the cloud will be mostly

transparent. Thus, far-infrared photons can carry out of the cloud some of the thermal energy generated by the collapse, and contraction can continue. Astronomers know this happens because they observe far-infrared spectral emission lines, called **cooling lines**, from dense cloud cores that are evidently contracting.

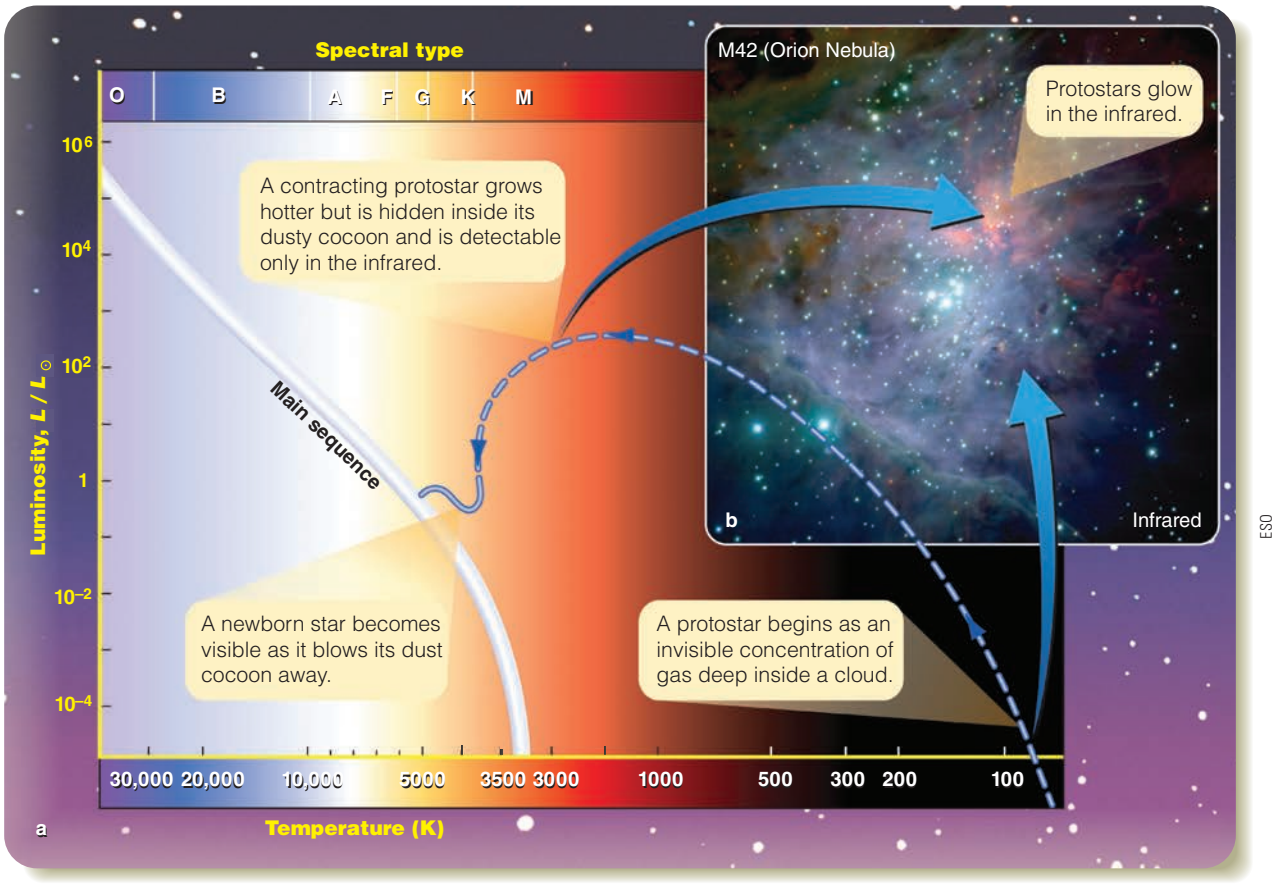
Your study of gas clouds has shown how the contraction of dense cores in giant molecular clouds can begin and how this contraction heats the gas. Now you are ready to construct a detailed story of the transformation from gas cloud to star.

11-1c Protostars

To understand star formation further, you can continue to imagine the contraction of a dense core as matter falls in and heats up, and the object gradually behaves more and more like a star. When the contracting core becomes hot enough, it produces short-wavelength radiation to which the core is not transparent. Then the core's thermal energy can no longer escape, and it enters a new stage of slow—rather than free-fall—contraction. Although the term is used rather loosely by astronomers, a **protostar** can be defined as a forming star that is compressed enough to be opaque at all wavelengths but not so hot and compressed as to be able to generate energy by nuclear fusion, as would a main-sequence star. Instead, a protostar shines with energy released gravitationally, from a combination of matter falling inward to the protostar plus slow contraction of the protostar itself.

Early in its life, a protostar develops a higher-density region at the center and a low-density outer region, or envelope, because the material that started falling in from closer to the center races ahead of material falling from farther out. Meanwhile, material continues to flow inward from the outer parts of the cloud. In other words, the cloud contracts from the inside out, with the protostar growing deep inside a surrounding cloud of cold, dusty gas.

These enveloping clouds have been called **cocoon nebulae** because they hide the forming protostar from view like a cocoon hiding a caterpillar becoming a moth. That is especially true for observations at short wavelengths. A cocoon nebula absorbs the protostar's radiation and, growing warm, reradiates the energy as infrared radiation with longer wavelengths than that produced by the protostar. This means that if you observe at short-infrared wavelengths, you may be able to observe the protostar itself; at longer wavelengths, you might be able to observe the much lower-temperature cocoon nebula that surrounds it. In either case, what you detect would be cooler than a star but much larger than the Sun and would therefore be a very luminous infrared source, with properties placing it beyond the upper right edge of the H–R diagram (Figure 11-4).



▲ **Figure 11-4** (a) This H–R diagram has been extended to very low temperatures to show schematically the contraction of a dim, cool protostar. (b) Protostars are normally invisible at visual wavelengths because they are cool objects and are deep inside dusty clouds of gas, but they are detectable at infrared wavelengths.

Theories and Proof

How do astronomers know the Sun isn't made of burning coal?

People say dismissively of a scientific explanation they dislike, "That's only a theory," as if a theory were just a random guess. A hypothesis is like a guess in some ways, although of course not a random one. What scientists mean by the word *theory* is a hypothesis that has "graduated" to being confidently considered a well-tested truth. You can think of a hypothesis as equivalent to having a suspect in a criminal case, whereas a theory is equivalent to finishing a trial and convicting someone of the crime.

Of course, no matter how many tests and experiments you conduct, you can never prove that any scientific theory is absolutely true. It is always possible that the next observation you make will disprove the theory. And it is unfortunately sometimes true that innocent people go to jail and guilty people are free, but, occasionally, with further evidence, those legal mistakes can be fixed.

There have always been hypotheses about why the Sun is hot. It was once thought that the Sun is a ball of burning material. Only a century ago, most astronomers accepted the hypothesis that the Sun is hot because gravity was making it contract. In the late 19th century, geologists showed that Earth was much older than the Sun could be if it was powered by gravity, so the "gravity hypothesis" had to be wrong. It wasn't until 1920 that a new hypothesis was proposed by Sir Arthur Eddington, who suggested that the Sun is powered somehow by the energy contained in atomic nuclei. In 1938 the German-American astrophysicist Hans Bethe showed how nuclear fusion could power the Sun. He won the Nobel Prize for that work in 1967.

The fusion hypothesis is now so completely confirmed that it is fair to call it a theory. No one will ever go to the center of the Sun, so you can't *prove* the fusion theory is right. Many observations and model calculations support this theory, and in Section 8-3d you saw further evidence in the neutrinos that have been detected coming from the Sun's core. Nevertheless there remains some tiny possibility that all the observations and models are misunderstood and that the theory will be overturned by some future discovery. Astronomers have tremendous confidence that the Sun is powered by fusion and not gravity or coal, but a scientific theory can never be proved conclusively correct.

There is a great difference between a theory in the colloquial sense of a far-fetched guess and a scientific theory that has undergone decades of testing and confirmation with observations, experiments, and models. However, no theory can ever be proved absolutely true. It is up to you as a consumer of knowledge and a responsible citizen to distinguish between a flimsy guess and a well-tested theory that deserves to be treated like truth—at least pending substantial further information.

ESA/NASA/SOHO/MDI

Technically it is still a theory, but astronomers have tremendous confidence that the Sun gets its power from nuclear fusion and not from burning coal.

The theory of star formation takes you into an unearthly realm filled with unfamiliar processes and objects. How can anyone really know how stars are born? Although the theory of star formation, like all scientific theories, can never be absolutely proven, it is far more than an initial hypothesis. Scientists have strong confidence in the theory of star formation because it has been repeatedly tested by observation (**How Do We Know? 11-1**).

11-2 The Orion Nebula: Evidence of Star Formation

You might ask, what evidence confirms the theory of star formation? Protostars are not easy to observe because they form deep inside dense interstellar clouds that block the view at visible wavelengths, and, being cooler than full-fledged stars, they produce radiation primarily at long wavelengths. Astronomers must therefore depend on observations at infrared and radio wavelengths to search for protostars in their natural environment. Furthermore, astronomers can estimate that the amount of time required for stars to form is only a few million years, which is a small fraction of their main-sequence lifetimes. Although this seems like a long time in human terms, it implies that almost all stars you observe will be in their much longer main-sequence stage and you won't be able to catch many stars in the relatively brief time they are protostars. That prediction, as you will learn, is confirmed by observations.

On a clear winter night, you can see with an unaided eye the Great Nebula of Orion—also known as Messier 42—as a fuzzy blob in Orion's sword. With binoculars or a small telescope it is striking, and through a large telescope it is breathtaking. At the center of the nebula lie four brilliant, blue-white stars known as the Trapezium, the brightest in a cluster of a few hundred stars. Surrounding the stars are the glowing filaments of a nebula more than 8 pc (26 ly) across. Like a great thundercloud illuminated from within, the churning currents of gas and dust suggest immense power. The significance of the Orion Nebula lies hidden around and beyond the visible nebula. This region is ripe with star formation.

11-2a Observing Star Formation

Observations provide plenty of evidence that star formation is a continuous process; you can be sure that stars are being born right now. Look at **Star Formation in the Orion Nebula**, and note four important points and one new term:

1. The visible nebula is only a small part of a vast, dusty molecular cloud. You see the nebula because the most massive stars born within it have ionized the gas and driven it outward, breaking out of the molecular cloud. Visual and infrared observations reveal small, dusty clouds of gas called *Bok globules* that may be forming into stars.
2. A single very hot and short-lived O star is responsible for producing most of the ultraviolet photons that ionize the gas and make the nebula glow. This star is so massive that it has already become a main-sequence star while its smaller siblings are still protostars.
3. Infrared observations reveal clear evidence of active star formation deeper in the molecular cloud behind the visible nebula.
4. Many of the young stars in the Orion Nebula are surrounded by disks of gas and dust.

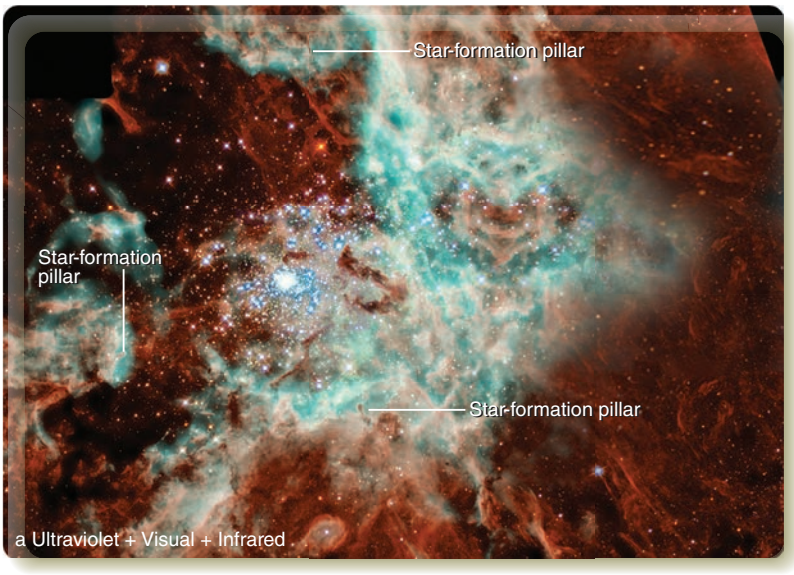
You should not be surprised to find star formation in Orion. The constellation is a brilliant pattern in the winter sky, marked by bright blue stars. Although those stars are hundreds of parsecs away, they appear bright because they are tremendously luminous. Such high-luminosity O and B stars cannot live more than a few million years. That is an astronomically short time, and astronomers know from measuring their Doppler shifts and proper motions that the O and B stars are moving relatively slowly through space. Therefore, they must have been born somewhere near where you see them now.

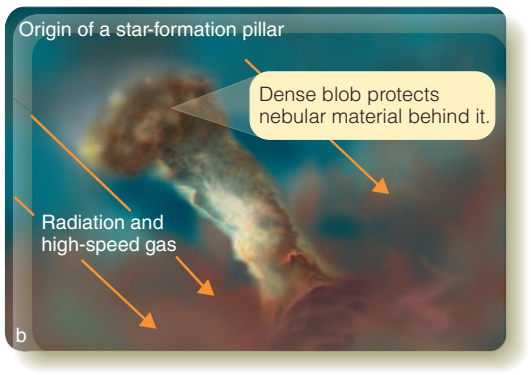
Are there other nebulae like the Orion Nebula? Such emission nebulae embedded in or connected to giant molecular clouds are actually quite common. Striking examples are NGC 2035 (right side of Figure 11-1) and NGC 346 (shown in the image that opens this chapter). Those emission nebulae, as well as others shown in this chapter and the previous chapter (for example, NGC 602 in Figure 10-10), are parts of star-forming regions produced by radiation from the most massive newborn stars just as in the Orion Nebula.

11-2b Contagious Star Formation

Astronomers have not only located firm evidence of star formation, they have also found evidence that star formation can stimulate more star formation. If a gas cloud produces massive stars, those massive stars ionize the gas nearby and drive it away. Where the intense radiation and hot gas pushes into surrounding gas, it can compress the gas and trigger more star formation. One sign of this process is the presence of **star-formation pillars**, which are columns of gas that point back toward the young, massive star. Such pillars are produced by denser regions of clouds that protect the gas behind them as the blast of intense radiation and hot gas flows past (**Figure 11-5**). Another way massive stars can trigger further star formation is by exploding as a supernova. As you read in the previous section, such an explosion can drive a shock wave through surrounding gas and cause interstellar clouds to collapse, forming new stars (Figure 11-3a).

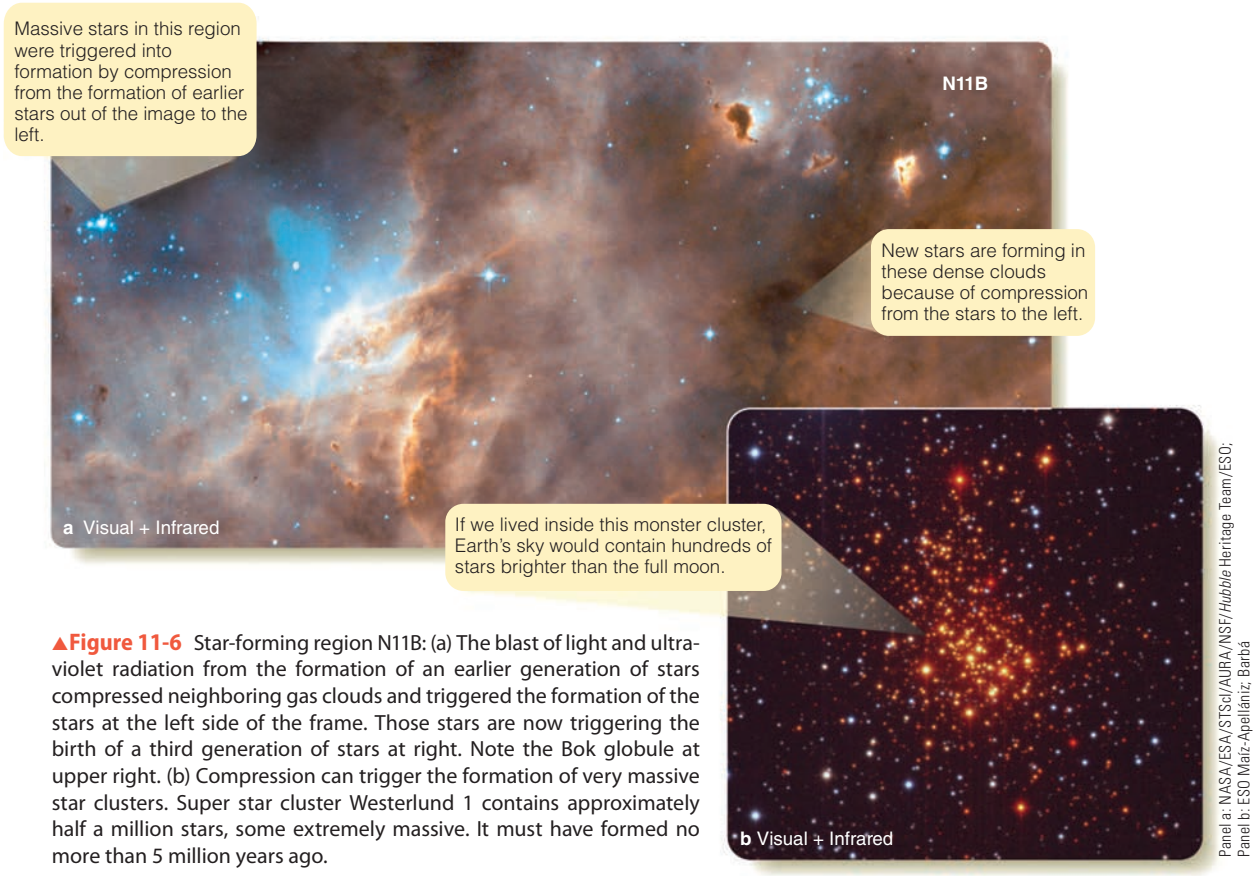
Because massive stars are effective in several ways at disturbing neighboring regions, they can cause star formation to spread through the ISM like a brush fire. Astronomers have located the remains of such episodes (**Figure 11-6**). Of course, lower-mass stars also form in the process, but they are much less able to trigger further star formation because they are not hot enough to ionize and drive away large amounts of gas, nor do they explode as supernovae.





Panel a: NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team; Panel b: ESO

Figure 11-5 (a) The hot, massive stars in the star cluster 30 Doradus are pouring out intense ultraviolet radiation and powerful winds of hot gas that are pushing back and compressing the surrounding nebula. (b) A close-up view shows slightly denser regions in the nebula protecting the gas behind them to make star-formation pillars a few light-years long, pointing back at the star cluster. The dense blobs are being compressed and could form more stars within the next few million years.


Massive stars in this region were triggered into formation by compression from the formation of earlier stars out of the image to the left.
N11B
New stars are forming in these dense clouds because of compression from the stars to the left.
a Visual + Infrared
If we lived inside this monster cluster, Earth's sky would contain hundreds of stars brighter than the full moon.
b Visual + Infrared
Panel a: NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/ESO;
Panel b: ESO Máiz-Apellániz; Barbá

▲**Figure 11-6** Star-forming region N11B: (a) The blast of light and ultraviolet radiation from the formation of an earlier generation of stars compressed neighboring gas clouds and triggered the formation of the stars at the left side of the frame. Those stars are now triggering the birth of a third generation of stars at right. Note the Bok globule at upper right. (b) Compression can trigger the formation of very massive star clusters. Super star cluster Westerlund 1 contains approximately half a million stars, some extremely massive. It must have formed no more than 5 million years ago.

The history of star formation in the constellation of Orion is written in its stars. Judging from the luminosities and estimated masses of the most prominent examples, the stars around Orion's west shoulder are about 12 million years old, whereas the stars of Orion's belt are clearly younger, no more than 8 million years old. The stars of the Trapezium at the center of the Great Nebula are only about 2 million years old. (For comparison, note that the Sun is 4600 million = 4.6 billion years old.) Star formation appears to have swept across Orion from northwest to southeast, beginning near Orion's west shoulder. The massive stars that formed there probably triggered the formation of the stars in Orion's belt. That star formation episode may in turn have caused the formation of the new stars you see today in the Great Nebula.

In the next million years, the familiar outline of the Great Nebula will change, and a new nebula may begin to form as the protostars embedded in a different part of the molecular cloud ionize the gas, drive it away, and become visible. Throughout the cloud, centers of star formation can be expected to develop and then dissipate as massive stars are born and force the gas to expand. If enough massive stars are born, they could blow the entire molecular cloud apart and bring the successive generations of star formation to a conclusion. The Great Nebula in Orion and its adjacent huge, but invisible, molecular cloud are a beautiful and dramatic example of the continuing cycles of star formation.

Practicing **SCIENCE**

What did Orion look like to the ancient Egyptians, to the first humans, and to the dinosaurs? Although we are small and short-lived compared with the stars, trying to reframe scientific concepts into human terms and scales sometimes helps scientists, and you, gain valuable perspective.

The Egyptian civilization had its beginning only a few thousand years ago, and that is not very long in terms of the history of Orion. The stars you see in the constellation are relatively young for stars, just a few million years old, so the Egyptians saw the same constellation you see. (They called it Osiris.) Even the Orion Nebula hasn't changed very much in a few thousand years, and Egyptians viewed it in the dark skies along the Nile and perhaps wondered about it.

Our oldest humanoid ancestors lived about 4 million years ago, and that was about the time that the youngest stars in Orion were forming. These earliest ancestors may have looked up and seen some of the stars you see, but other stars have formed since that time. The Great Nebula's emission is excited by the Trapezium stars that are no more than about 2 million years old, so our earliest humanoid relatives probably didn't see the Great Nebula.

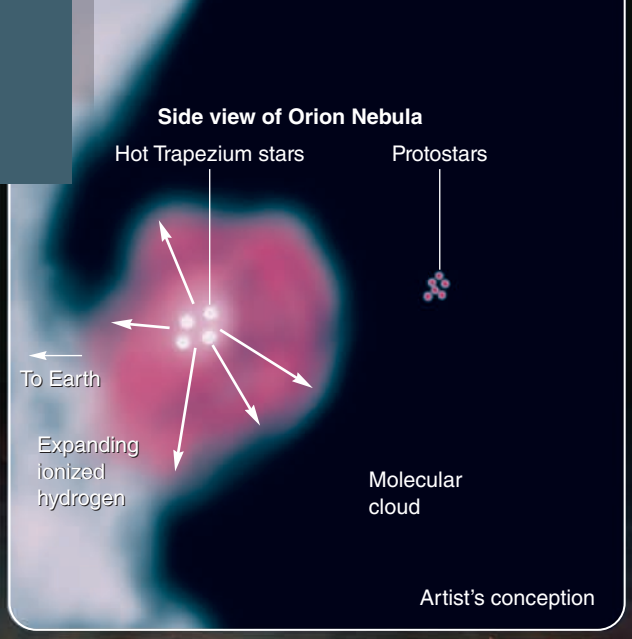
The dinosaurs would have seen something quite different. The last of the dinosaurs died about 65 million years ago, long before the birth of the brightest stars in Orion. The dinosaurs, had they the brains to appreciate the view, might have seen bright stars along the Milky Way, but they didn't see Orion. All of the stars in the sky are moving through space, and the Sun is orbiting the center of our Galaxy. Over many millions of years, the stars move appreciable distances across the sky. The night sky above the dinosaurs contained totally different star patterns.

Star Formation in the Orion Nebula

1 The visible Orion Nebula shown below is a pocket of ionized gas on the near side of a vast, dusty molecular cloud that fills much of the southern part of the constellation Orion. The molecular cloud can be mapped by radio telescopes. To scale, the cloud would be many times larger than this page. As the stars of the Trapezium were born in the cloud, their radiation has ionized the gas and pushed it away. Where the expanding nebula pushes into the larger molecular cloud, it is compressing the gas (see diagram at right) and may be triggering the formation of the protostars that can be detected at infrared wavelengths within the molecular cloud.

Hundreds of stars lie within the nebula, but only the four brightest, those in the Trapezium, are easy to see with a small telescope. A fifth star, at the narrow end of the Trapezium, can be visible on nights of good seeing.

The cluster of stars in the nebula is less than 2 million years old. This means the nebula is similarly young.



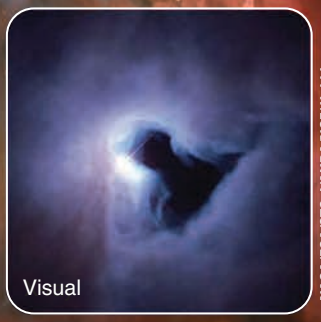
NASA/JPL-Caltech/N. Flagey (IAS), A. Noriega-Crespo (SSC)

The near-infrared image above reveals more than 50 low-mass, cool protostars.

Trapezium

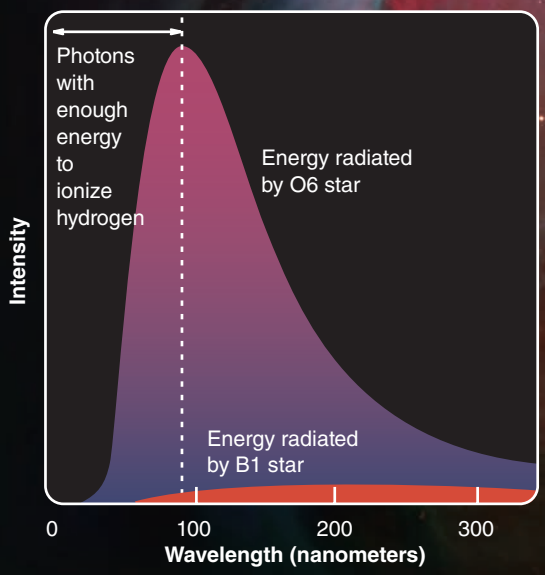
Visual

NASA/ESA/STScI/AURA/NSF/M. Robberto/*Hubble Space Telescope* Orion Treasury Project Team



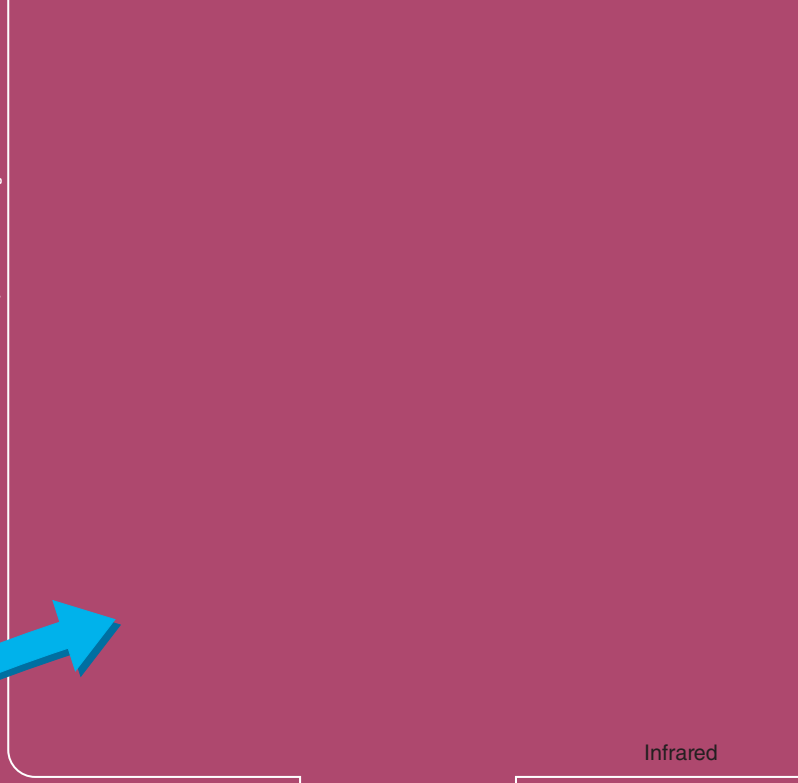
NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

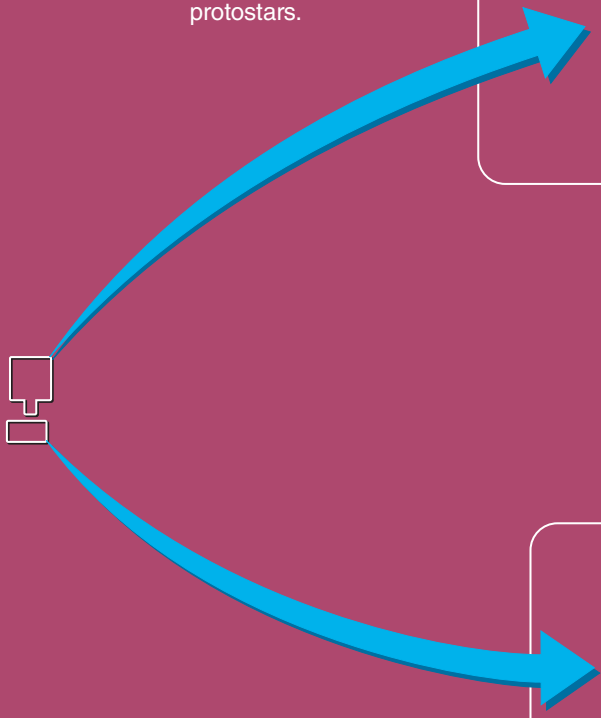
Small dark clouds called **Bok globules**, named after astronomer Bart Bok, are found in and near star forming regions. The one pictured here is part of nebula NGC 1999 near the Orion Nebula. Typically about 1 ly in diameter, they contain from 10 to 1000 solar masses.

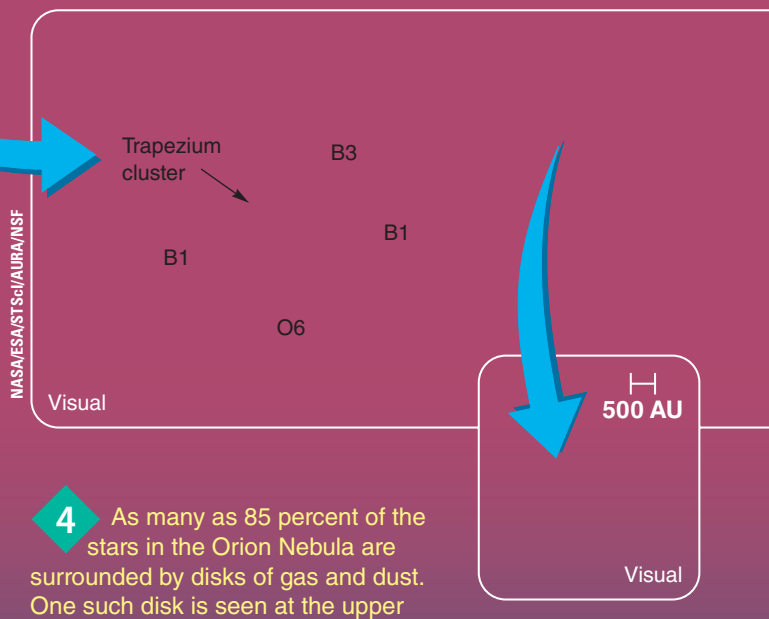


2 Of all the stars in the Orion Nebula, just one is hot enough to ionize the gas. Only photons with wavelengths shorter than 91.2 nm can ionize hydrogen. The second-hottest stars in the nebula are B1 stars, and they emit little of this ionizing radiation. The hottest star, however, is an O6 star that has 30 times the mass of the Sun. At a temperature of 40,000 K, it emits plenty of photons with wavelengths short enough to ionize hydrogen. Remove that one star, and the nebula's emission would turn off.

3 The infrared image from the *Spitzer Space Telescope* reveals extensive nebulosity surrounding the visible Orion Nebula. Red and orange show the locations of warm dust that has been heated by starlight, green shows hot dust and ionized gas, and blue shows light coming directly from stars.


AAO/David Malin Images
In this near-infrared image, fingers of gas (nicknamed the "Hand of God") rush away from of a region full of infrared protostars.
Infrared
NASA/IRTF/D. Gezari (GSFC), D. Backman (Franklin & Marshall College), M. Werner (JPL-Caltech)
Infrared
BN
KL
The Becklin-Neugebauer (BN) object is a B star just reaching the main sequence. The Kleinmann-Low (KL) Nebula is a cluster of cool, young protostars.
BN and KL are detectable only at infrared wavelengths.





The spectral types of the Trapezium stars are shown here. The gas looks green because of filters used to record the image.

4 As many as 85 percent of the stars in the Orion Nebula are surrounded by disks of gas and dust. One such disk is seen at the upper right of this *Hubble Space Telescope* image, magnified in the inset. Radiation from the nearby hot, luminous Trapezium stars is evaporating gas from the disk and driving it away to form an elongated nebula.

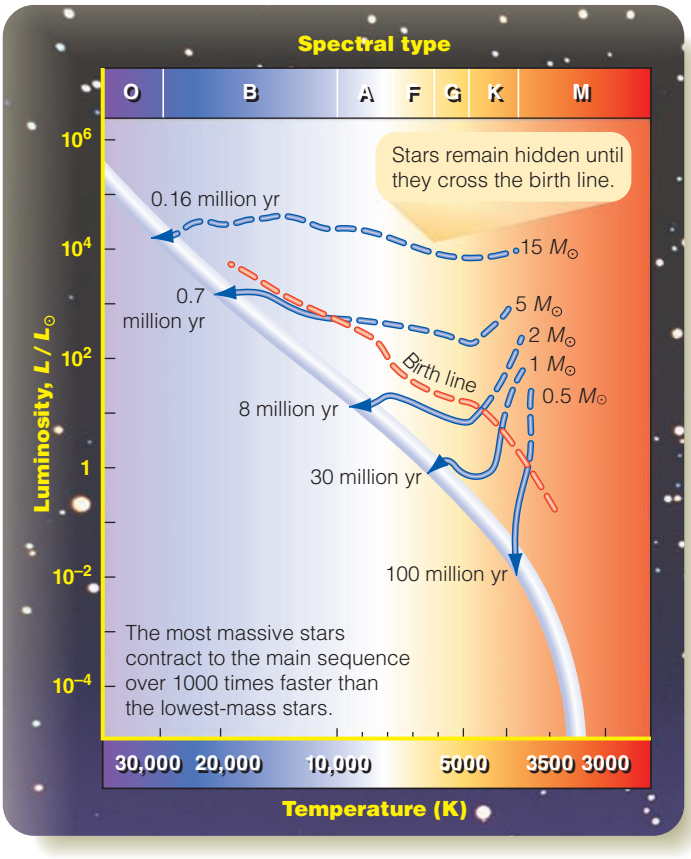
Infrared

11-3 Young Stellar Objects and Protostellar Disks

The more massive a star, the stronger its gravity and therefore the more rapid will be its contraction. Astronomers calculate that the Sun took about 30 million years from when its cloud began contracting until it became a main-sequence star, but a 30-solar-mass star would take only 30,000 years. A 0.2-solar-mass star needs almost 1 billion years to finish forming and reach the main sequence.

11-3a Young Stellar Objects

When most of the material in a protostar's cocoon nebula has fallen inward or been driven away, the protostar will no longer be quite so hidden. The location in the H–R diagram of protostars that have become detectable at visible wavelengths because their cocoons have disappeared is called the **birth line** (Figure 11-7). Once a star crosses the birth line and becomes visible, it continues to contract and move toward the main sequence at a pace that depends on its mass. Stars in this late stage of formation are sometimes called **Young Stellar Objects (YSOs)** or pre-main-sequence stars, to distinguish them from earlier protostellar stages.

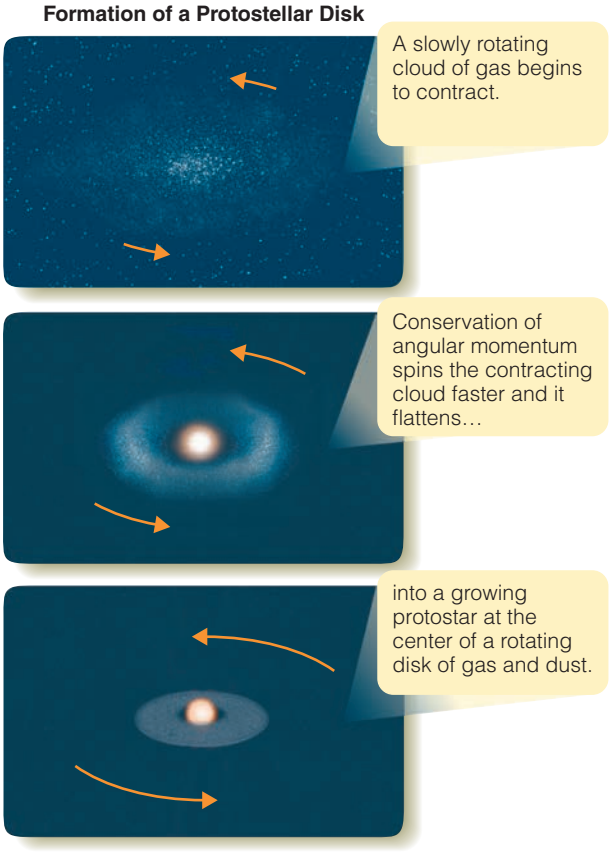


▲ **Figure 11-7** The more massive a protostar is, the faster it contracts. A 1-solar-mass star requires 30 million years to reach the main sequence. The dashed line is the birth line, where contracting protostars first become visible as they dissipate their surrounding clouds of gas and dust. Compare with Figure 11-4, which shows the evolution of a protostar of about 1 solar mass as a dashed line up to the birth line and as a solid line from the birth line to the main sequence.

11-3b Protostellar Disks

As a contracting molecular cloud core becomes a protostar, the cloud rotates faster and faster. The rapidly spinning core of the cloud must flatten into a spinning disk like a blob of pizza dough spun into the air. Gas from the envelope that had little angular momentum to begin with, or lost angular momentum through collisions, can sink directly to the center of the cloud, where the protostar grows larger, surrounded by the disk. Other infalling material can add to the disk and then move within the disk plane toward the star. Astronomers calculate that this material gives up much of its angular momentum via collisions in the disk and possibly also by interactions with the disk's magnetic field, before sinking into the protostar (Figure 11-8).

Theoretically, for a protostar to have no surrounding disk at all, its original cloud would need to have zero angular momentum, which is very unlikely. Astronomers studying protostars find that the majority of them appear to be surrounded by disks of gas and dust. The disks that form around protostars are called **protostellar disks**, and they are important because, as you will learn in a later chapter, astronomers conclude that planets form within these disks. In fact, the evidence indicates that Earth


Formation of a Protostellar Disk
A slowly rotating cloud of gas begins to contract.
Conservation of angular momentum spins the contracting cloud faster and it flattens…
into a growing protostar at the center of a rotating disk of gas and dust.

▲ **Figure 11-8** The rotation of a contracting gas cloud forces it to flatten into a disk, and the protostar grows at the center. The scale of the top panel is much larger than that of the lower two panels.

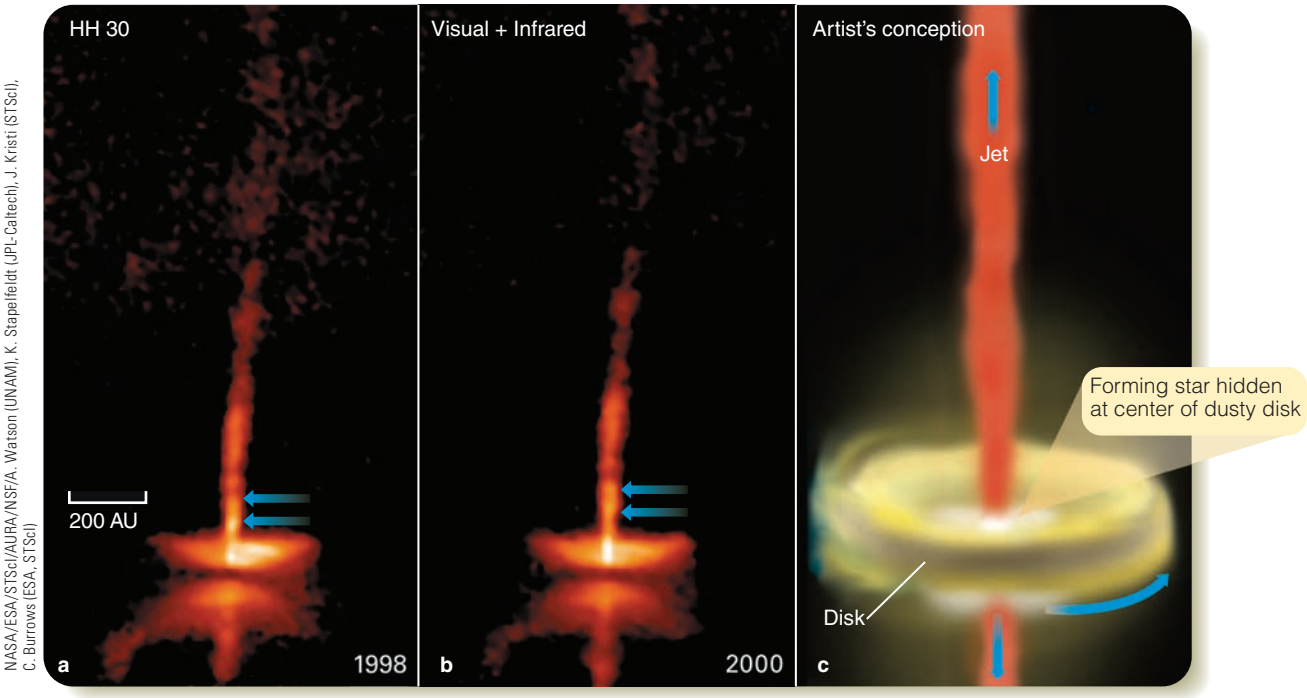
formed in such a disk around the "protosun" 4.6 billion years ago. Because most protostars have protostellar disks, it seems likely that most stars have planetary systems. You will learn more in a later chapter about planets orbiting stars other than the Sun.

Look through **Observations of Young Stellar Objects and Protostellar Disks** and notice that it makes four important points and introduces three new terms:

1. Star formation regions contain types of stars so young they must have formed recently. *T Tauri* stars, for example, are understood to be relatively low-mass protostars, ranging from roughly 0.5 to 2 solar masses, still in a slowly contracting pre-main-sequence phase.
2. Infrared observations reveal the complicated shapes of star-forming clouds that are strongly affected by radiation and winds from nearby massive, luminous stars. The formation of massive stars may both foster and disturb the formation of their lower-mass siblings.
3. In the H–R diagram, newborn stars lie between the birth line and the main sequence—with just the properties you would expect of stars that have recently lost their dust cocoons. Young stars tend to have very active, hot coronae and chromospheres, making them prominent X-ray sources.
4. Observations provide clues about the effects of disks of gas and dust around protostars. These disks evidently cause *bipolar flows* that push into the surrounding ISM and produce glowing blobs called *Herbig–Haro* (pronounced *Herbig-Airo*) *objects*. Astronomers also have abundant evidence that these disks are the sites of planet formation.

In some cases, these dark disks of gas and dust are clearly visible around newborn stars (**Figure 11-9**), but it isn't clear how protostellar disks produce jets. Certainly, such large, contracting, spinning disks contain tremendous energy. Theorists suspect that magnetic fields become twisted tightly around the disks. Exactly how those fields eject hot gas away from the stars is not completely understood, but the presence of confining disks is thought to explain why material flows out in two streams along their respective axes of rotation. The detection of these jets was one of the first pieces of evidence that substantial disks often surround protostars.

Observations and models of protostellar disks indicate that most of the disk material eventually flows into the star and adds to its mass, emitting thermal radiation as it hits the protostar's surface. In fact, a large fraction of the luminosity of some protostars seems to come from this accretion flow. The flow is not completely steady but includes episodes during which Jupiter-mass blobs of gas and dust suddenly fall into the protostar from the disk, causing temporary outbursts of increased luminosity. This pattern of accretion surges and resulting flares is called the **FU Orionis phenomenon** (or FU Ori, for short) after the first star in which it was identified. Protostellar disks and FU Ori outbursts were first identified around protostars with approximately the mass of the Sun, but further research indicates that disks and FU Ori flares are part of the story of high-mass star formation as well.



NASA/ESA/STScI/AURA/NSF/A. Watson (UNAM), K. Stapelfeldt (JPL-Caltech), J. Kristi (STScI), C. Burrows (ESA, STScI)

▲**Figure 11-9** (a) Newly formed star HH 30 lies at the center of a dense disk of dusty gas that is narrow near the star and thicker farther away. Although the star is hidden by the edge-on dusty disk, the star illuminates the inner surface of the disk. Interactions between infalling material in the disk and the spinning star eject bipolar jets of gas along the axis of rotation. (The images are cropped to emphasize the upper jet.) (b) Arrows indicate knots of material that have moved noticeably in two years. (c) The artist's conception can help interpret what can be seen in the *Hubble Space Telescope* images in a and b.

Observations of Young Stellar Objects and Protostellar Disks

1 The nebula around the star S Monocerotis (constellation of the Unicorn) is bright with hot stars. Such stars live short lives of only a few million years, so they must have formed recently. Such regions of young stars are common. The entire constellation of Orion is filled with young stars and clouds of gas and dust.

Nebulae containing young stars usually contain **T Tauri stars**. These stars fluctuate irregularly in brightness, and many are bright in the infrared, indicating that they are surrounded by dust clouds and in some cases by dust disks. Doppler shifts show that gas is flowing away from many T Tauri stars. The T Tauri stars are evidently protostars and young stars just blowing away their dust cocoons. They have estimated ages ranging from 100 thousand to 100 million years. Spectra of T Tauri stars show signs of an active chromosphere as you might expect from young, rapidly rotating stars with powerful dynamos and strong magnetic fields.

Skyfactory.org

NASA/JPL-Caltech/W. Reach

2 The Elephant Trunk (*above*) is a dark nebula compressed and twisted by radiation and winds from a luminous star beyond the left edge of this image. Infrared observations reveal that it contains six protostars (*pink objects, lower edge*) not detected at visual wavelengths.

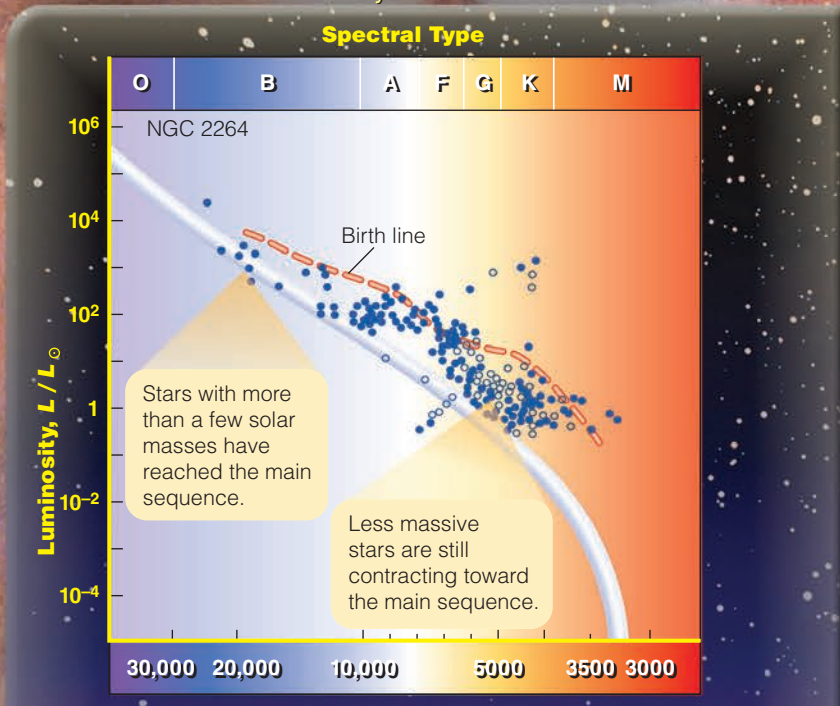
NASA/*Hubble* Heritage Project

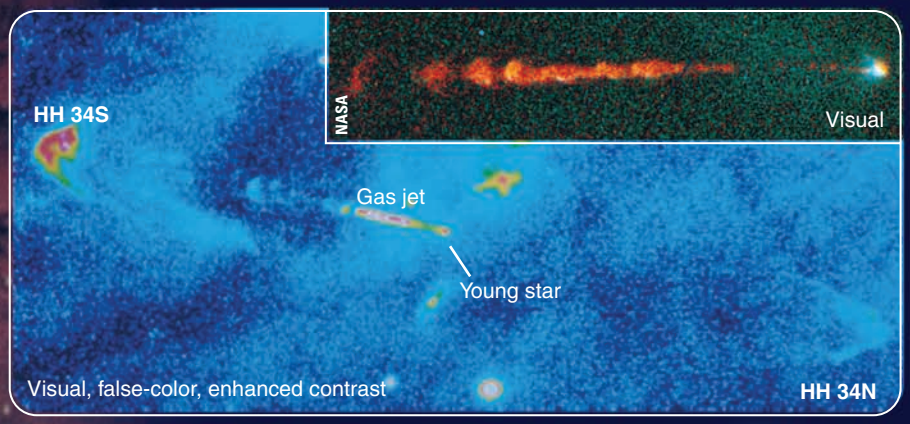
Visual

NASA/Penn State University

Approximately 1000 young stars with hot chromospheres appear in this X-ray image of the Orion Nebula.

3 The star cluster NGC 2264, embedded in the nebula pictured on this page, is only a few million years old. Lower-mass stars have not yet reached the main sequence, and the cluster contains many T Tauri stars (open circles), which have properties plotted near the birth line, above (higher luminosity), and to the right (cooler) than the main-sequence stars. The least luminous stars in the cluster were too faint to be observed in this study.

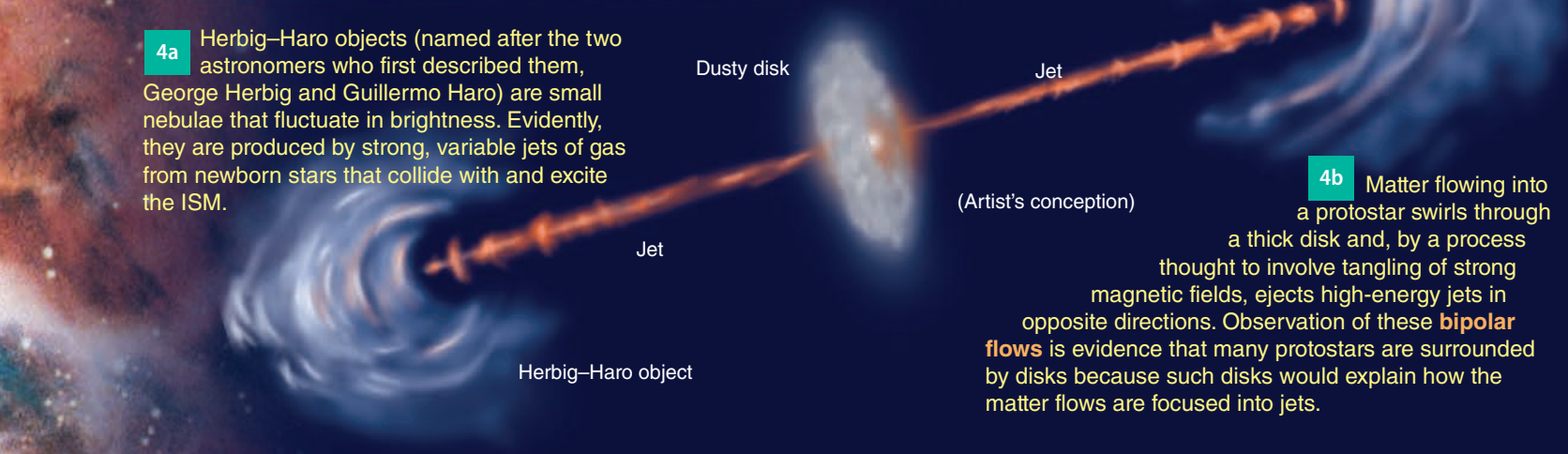




Credit for main panel: Calar Alto/MPIA/IAA/R.Mundt; Credit for inset: NASA/ESA/STScI/AURA/NSF

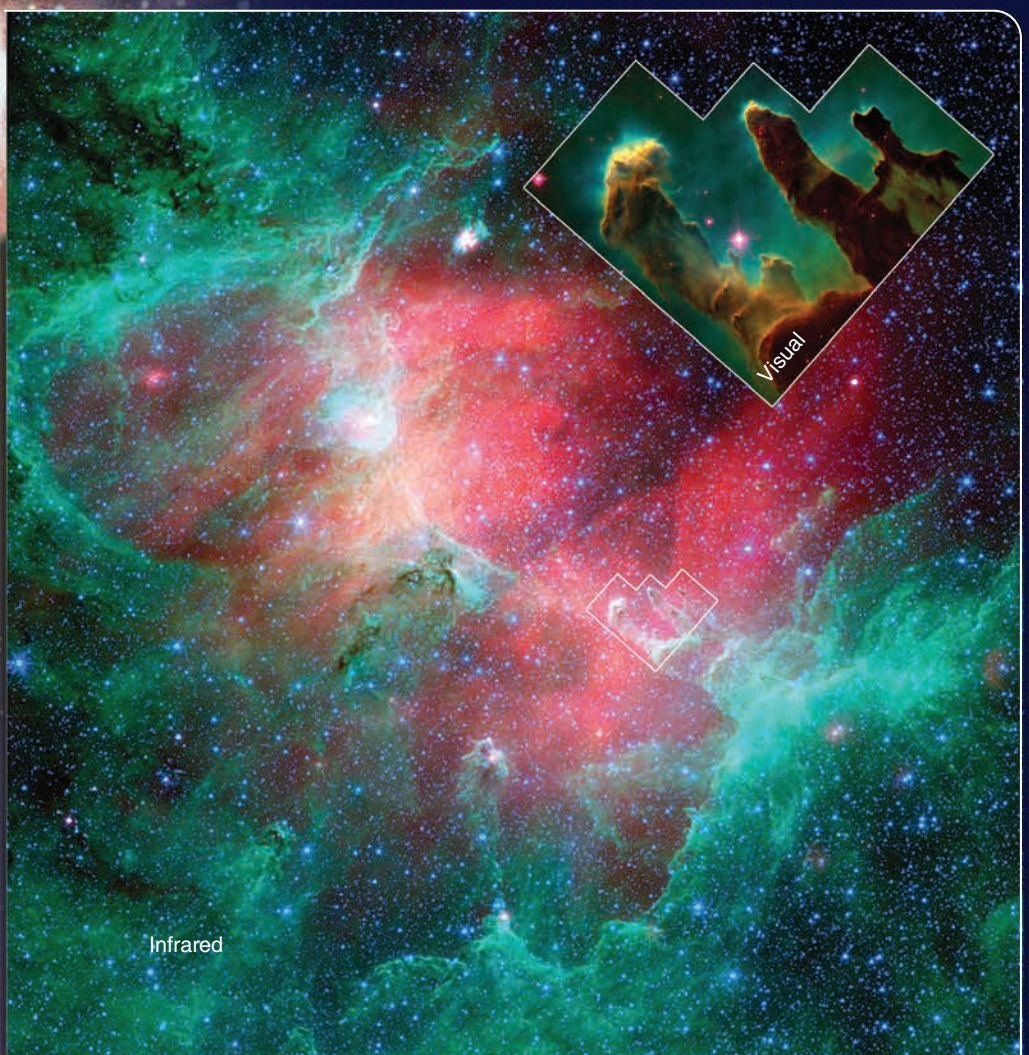
4 A newborn star shown at the center of this image is emitting powerful jets to left and right. Where the jets strike the ISM they produce **Herbig–Haro objects**. The inset shows how irregular the jet is. Such jets can be more than a light-year long and contain gas traveling at 100 km/s or more.

4a Herbig–Haro objects (named after the two astronomers who first described them, George Herbig and Guillermo Haro) are small nebulae that fluctuate in brightness. Evidently, they are produced by strong, variable jets of gas from newborn stars that collide with and excite the ISM.



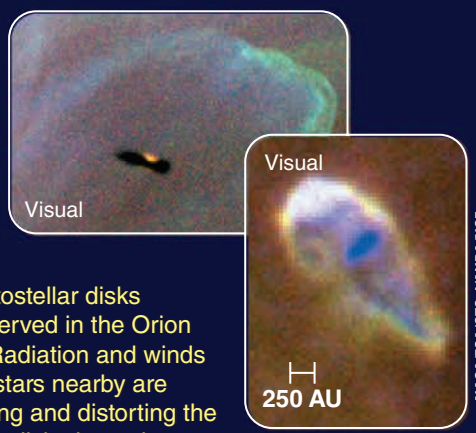
(Artist's conception)

4b Matter flowing into a protostar swirls through a thick disk and, by a process thought to involve tangling of strong magnetic fields, ejects high-energy jets in opposite directions. Observation of these **bipolar flows** is evidence that many protostars are surrounded by disks because such disks would explain how the matter flows are focused into jets.



Credit for main panel: NASA/JPL-Caltech/N. Flagey (IAS), A. Noriega-Crespo (SSC); Credit for inset:NASA/ESA/STScI/AURA/NSF/J. Hester & P. Scowen

4c Radiation and winds from massive stars have shaped this nebula, Messier 16, and a recent supernova has heated some of the dust (*red*). Shock waves from the explosion will destroy the Eagle Nebula star-formation pillars (*inset*) within about 1000 years. Erosion of part of the Eagle Nebula has exposed small fingers and globules of denser gas. About 15 percent of those globules have formed protostars.



NASA/ESA/STScI/AURA/NSF

4d Protostellar disks observed in the Orion Nebula: Radiation and winds from hot stars nearby are evaporating and distorting the disks. The disk above is seen silhouetted against the nebula, with light from the protostar that is embedded in the disk's center apparent at the upper edge. Although substantially larger than the present size of our Solar System,

fo

When a protostar becomes hot enough, it can drive away the gas and dust of the protostellar disk as well as any remaining traces of its cocoon. Just as the Sun exhales a solar wind, stars produce **stellar winds**, which can be quite vigorous for hot, young stars. Also, when photons encounter gas atoms or dust grains in space, the photons can exert **radiation pressure**. Stellar winds and radiation pressure combine to blow disks and remnant cocoons apart. As you learned in the previous section, the most luminous stars are even capable of eroding the cocoon and disk material of their neighbors (Figure 11-5). You can recognize that process in the overall shape of the Eagle Nebula (Figure 4c of the Concept Art: Observations of Young Stellar Objects and Protostellar Disks) as well as in the smaller pillars within that nebula.

11-3c Young Stellar Associations

Some evidence of stellar youth is subtle. An **association** is a widely distributed star cluster that is not dense enough to be permanently held together by its own gravity; its stars wander away as time goes on. It is not clear why some gas clouds give birth to compact star clusters held together by their own gravity, whereas others give birth to larger associations that are not bound together. You can safely conclude, however, that associations must consist of young stars because the stars drift apart quickly, astronomically speaking. The constellation Orion, a known region of star formation, is filled with low-mass T Tauri protostars in a **T association**. **OB associations**, fairly obviously, are extended groups of more massive O and B stars. Associations of stars that are journeying away from each other as we watch are clear evidence of recent star formation. Careful observations with large telescopes and sensitive infrared detectors indicate that T and OB associations include not only stars, but also numerous lower-mass brown dwarfs and even free-floating, solitary Jupiter-mass planets that formed along with the stars.

Practicing SCIENCE

What evidence can you cite that stars in so-called star-forming regions are actually young? Science is based on evidence, and there is no point to building a theory of star formation unless it is standing on a foundation of observational evidence.

First, you should note that some extremely luminous stars can't live very long because they must completely use their fuel supplies in an astronomically short time. So, when you see stars such as the hot, luminous O and B stars in Orion, you know they must have formed in the past few million years. Also, many regions of gas and dust contain bright, hot stars that are caught in the act of blowing apart the nebulae in which they are embedded; those stars, too, must have formed recently or those nebulae would be gone already.

You have other evidence when you look at T Tauri stars. They: (1) are usually associated with interstellar gas and dust concentrations, (2) have H–R diagram locations just above the main sequence where you expect stars to be that are still contracting, and (3) often have protostellar disks that are easily destroyed and cannot last long. There seems to be no doubt that star formation is an ongoing process.

Now, consider the idea that planets such as Earth form as a by-product of star formation. ***What evidence can you cite that protostars are often surrounded by disks of gas and dust?***

11-4 Stellar Structure

How would you go about making a star? It's simple, really. All you need is something—maybe a light-year-sized broom—to gather together a star's mass of ISM, and gravity will do the rest. The object that forms in that way is also conceptually simple.

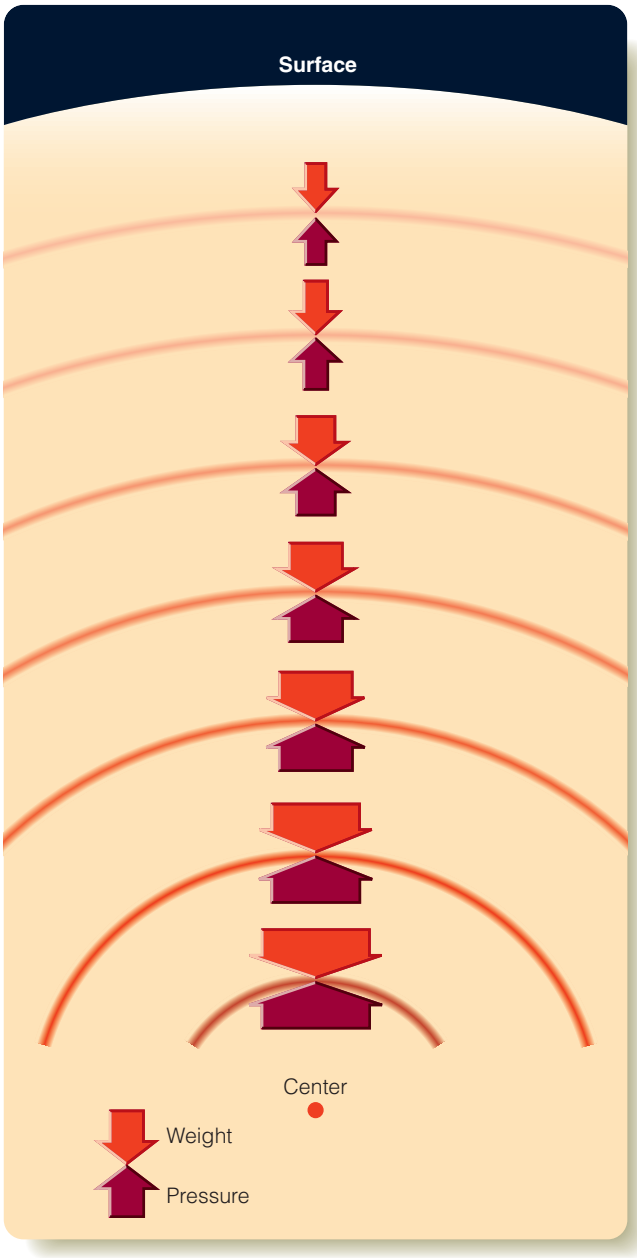
11-4a What Keeps a Star Stable?

If there is a single idea in stellar astronomy that can be called crucial, it is the concept of balance. In this section, you will discover that stars are held together by their own gravity balanced by the support of their internal heat and pressure. This story will lead your imagination into a region where you yourself can never go—the heart of a star.

Consider the structure of a star, using the basic concept of balance. What is meant here by structure is the variation in temperature, density, pressure, composition, and so on, between the surface of the star and its center. You can think more easily about stellar structure if you imagine that the star is divided into many concentric layers like those in an onion. You can then focus on the temperature, pressure, and density in each layer. These helpful layers exist only in the imagination; stars generally do not have such truly distinguishable layers.

The weight of each layer of a star must be supported by the layer below. (Remember, the words *down* or *below* are conventionally used to refer to regions closer to the center of a star.) Picture a pyramid of people in a circus stunt: The people in the top row do not have to hold up anybody else; the people in the next row down are holding up the people in the top row, and so on. In a star that is stable, the deeper layers must support the weight of all of the layers above. The inside of a star is made up of gas, so the weight pressing down on a layer must be balanced by gas pressure in that layer. If the pressure is too low, the weight from above will compress and push down the layer, and if the pressure is too high, the layer will expand and lift the layers above.

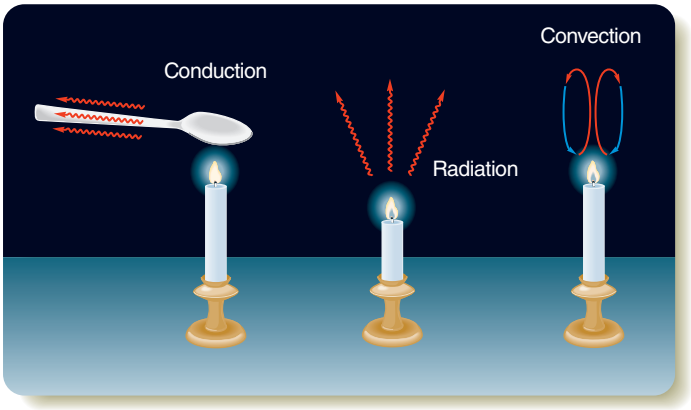
This balance between weight and pressure is called **hydrostatic equilibrium**. *Hydro* (from the Greek word for water) tells you the material is a fluid, which by definition includes gases as in a star, and *static* tells you the fluid is stable, neither expanding nor contracting. Figure 11-10 shows this hydrostatic balance in the imaginary layers of a star. The weight pressing down on each layer is shown by lighter red arrows, which grow larger with increasing depth to represent the weight growing larger. The pressure in each layer is shown by darker red arrows. The important point here is that you can be absolutely sure, from this simple argument, that the pressure inside a stable star must grow larger with increasing depth to support the weight and keep the star stable, even if you can't directly examine the inside of a star.



▲ **Figure 11-10** The law of hydrostatic equilibrium says the pressure in each layer must balance the weight on that layer. Consequently, as the weight increases from the surface of a star to its center, the pressure must also increase.

The pressure in a gas depends on the temperature and density of the gas. Near the surface there is not much weight pressing down, so the pressure does not need to be high for stability. Deeper in the star, the pressure must be higher, which means that the temperature and density of the gas must also be higher. In other words, the principle of hydrostatic equilibrium tells you that stars must have not only high pressure but also high temperature and density inside to support their own weight and be stable.

Although the law of hydrostatic equilibrium can tell you some things about the inner structure of stars, you also need to know how energy flows within the star to completely understand its structure.



▲ **Figure 11-11** The three modes by which energy may be transported from the flame of a candle, as shown here, are the three modes of energy transport within a star.

11-4b Energy Transport

The surface of a star radiates light and heat into space and would quickly cool if that energy were not replaced. Because the inside of the star is hotter than the surface, energy must flow outward to the surface, where it radiates away. This flow of energy through each layer determines its temperature, which, as you saw previously, determines how much weight that layer can balance. The movement of energy from the inside to the surface of a star thus plays a crucial role in determining the star's structure.

The law of **energy transport** says that energy must flow from hot regions to cooler regions by either conduction, convection, or radiation (**Figure 11-11**). Conduction is the most familiar form of heat flow. If you hold the bowl of a spoon in a candle flame, the handle of the spoon grows warmer as heat, in the form of motion among the atoms of the spoon, is conducted from atom to atom up the handle. Note that conduction is not a significant cause of heat flow in normal stars because radiation or convection are much more efficient, even in their centers; conduction is only important in rare types of stars with extremely high densities.

The transport of energy by radiation is another familiar experience. Put your hand near a candle flame, and you can feel the heat. What you actually feel are infrared photons—packets of energy—radiated by the flame and absorbed by your hand. Radiation is the principal means of energy transport in the interiors of most types of stars. Photons are absorbed and reemitted in random directions over and over as energy works its way from the hot interior toward the cooler surface. The flow of energy by radiation is controlled by the **opacity** of the gas; in other words, its resistance to movement of radiation. Opacity in turn depends strongly on the temperature: A hot gas is less opaque, more transparent, than a cool gas.

If the opacity is high, radiation cannot flow through the gas easily, and it backs up like water behind a dam. When enough heat builds up, the gas begins to churn as hot gas rises upward and cool gas sinks downward. This heat-driven circulation of a fluid is called *convection*, which is the third way energy can move

inside a star. You are familiar with convection: The rising wisp of smoke above a candle flame is carried by convection. Energy is carried upward in these convection currents as rising hot gas (red in the right-hand diagram within Figure 11-11) and also as sinking cool gas (blue in the right-hand diagram within Figure 11-11). Convection is important not only because it carries energy but also because it mixes the gas.

11-5 The Source of Stellar Energy

In the previous section, you learned that the Sun and other stars are stable because their centers have very high temperature and, therefore, pressure. Stars are born when gravity pulls interstellar matter together. When the density and temperature at the centers of the new stars become high enough, nuclear fusion begins making energy. This supplies the energy that keeps the core hot enough, with high enough pressure, that the star can be stable. In this section, you will learn the details of how that happens.

11-5a A Review of the Proton–Proton Chain

In Section 8-3, you studied the center of the Sun and discovered that it creates energy through hydrogen fusion in a series of nuclear reactions called the proton–proton chain (look back to Figure 8-16). That reaction begins with the fusion of two protons, which are hydrogen nuclei. Protons have positive charges and repel each other electrically. As you learned in Section 8-3b, electrical force is also called the Coulomb force, so the resistance of protons to being combined is called the Coulomb barrier. High-speed collisions are required to penetrate the Coulomb barrier, and high speed for atoms and subatomic particles means high temperature. Consequently, the proton–proton chain cannot effectively occur if the gas temperature is lower than about 4 million K.

You also discovered that the gas must be dense if the proton–proton chain is to produce significant energy. The fusion of two protons is unlikely, even if they collide, so a huge number of collisions are necessary to produce a few fusion reactions. Furthermore, a single pass through the proton–proton chain produces only a tiny amount of energy, so a vast number of reactions are needed to supply enough energy to support a star. The higher the gas density for a given temperature, the larger will be the number of fusion reactions.

Those are the reasons why the proton–proton chain produces energy only near the Sun's center, where the temperature and density are high. In fact, only about the innermost 30 percent of the Sun's radius has the right conditions to support fusion. The rest of the Sun, farther from the center, isn't hot and dense enough.

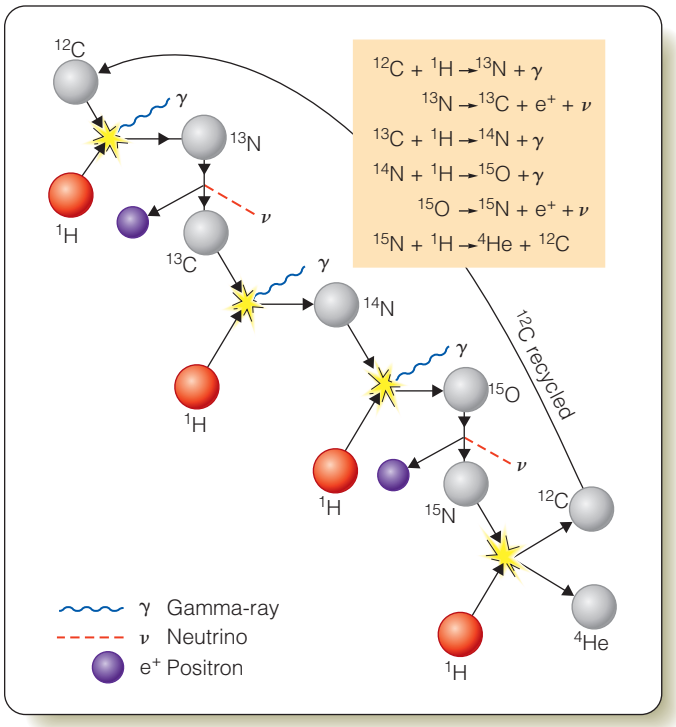
You might expect other stars to fuse hydrogen the same way the Sun does, and you would be right for most stars. Some stars, however, fuse hydrogen using a different reaction, and that makes a big difference to their structure and lifetimes.

11-5b The CNO Cycle

Models of main-sequence stars more massive than 1.1 solar masses indicate they are hot enough to fuse hydrogen into helium using the **CNO cycle**, which is a hydrogen fusion process that uses carbon, nitrogen, and oxygen as stepping-stones.

Look carefully at **Figure 11-12** and notice the CNO cycle sequence. The cycle begins with a carbon-12 nucleus absorbing a proton and becoming nitrogen-13, which decays to become carbon-13. The carbon-13 nucleus absorbs a second proton and becomes nitrogen-14, which absorbs a third proton and becomes oxygen-15. The oxygen-15 decays to become nitrogen-15, which absorbs a fourth proton, ejects a helium nucleus, and becomes carbon-12. The net result is four protons combining to make a helium nucleus, the same as in the proton–proton chain. Notice that carbon-12 begins the CNO cycle and ends the cycle, so the carbon-12 nuclei can be recycled over and over. This CNO cycle has the same outcome as the proton–proton chain, but it is different in an important way.

The CNO cycle begins with a carbon nucleus combining with a proton. Because a carbon nucleus has a positive charge six times higher than a proton, the Coulomb barrier is higher for this reaction than for the proton–proton chain. Temperatures


$^{12}C + {}^{1}H \rightarrow {}^{13}N + \gamma$
$^{13}N \rightarrow {}^{13}C + e^{+} + \nu$
$^{13}C + {}^{1}H \rightarrow {}^{14}N + \gamma$
$^{14}N + {}^{1}H \rightarrow {}^{15}O + \gamma$
$^{15}O \rightarrow {}^{15}N + e^{+} + \nu$
$^{15}N + {}^{1}H \rightarrow {}^{4}He + {}^{12}C$
^{12}C recycled
γ Gamma-ray
ν Neutrino
e^{+} Positron

▲**Figure 11-12** The CNO cycle uses carbon-12 (^{12}C) as a catalyst to combine four hydrogen nuclei (^{1}H) to make one helium nucleus (^{4}He) plus energy. The carbon nucleus reappears at the end of the process, ready to start the cycle over.

higher than 16 million K are required so that significant numbers of protons can penetrate the Coulomb barriers of carbon nuclei, much higher than the proton–proton chain's ignition temperature of 4 million K. The CNO cycle therefore needs much hotter gas to operate than does the proton–proton chain. The temperature at the very center of the Sun is estimated to be just under 16 million K, which is why the Sun makes nearly all of its energy from the proton–proton chain and only a little from the CNO cycle. The high temperature threshold for the CNO cycle means that main-sequence stars with the hottest cores, O and B types, create nearly all of their energy via the CNO cycle. Nevertheless, the CNO cycle and the proton–proton chain have the same net result of fusing hydrogen to make helium.

11-5c Inside Stars

There is a **Common Misconception** that the Sun's core is hot because nuclear reactions occur there. That statement reverses cause and effect. Rather, it is correct to say the Sun and other stars have nuclear reactions in their cores because the temperature is hot enough there. You observe the Sun to be stable; therefore, it must have a certain temperature, pressure, and density in its core. It would have about that same temperature in its core even if somehow you could magically turn off the nuclear reactions. But then, without a source of energy to replace the luminosity pouring out into space, the Sun would gradually lose its internal energy and start slowly contracting again, as was true during its protostellar phase before the nuclear reactions started.

In Section 9-6c, you discovered that the stars on the main sequence are ordered according to mass. Combining this with what you know about hydrogen fusion reveals that there are two kinds of main-sequence stars: (1) massive stars on the upper main sequence fuse hydrogen into helium via the CNO cycle, in which carbon, nitrogen, and oxygen nuclei act as catalysts but are not consumed; (2) low-mass stars on the lower main sequence, including the Sun, fuse hydrogen into helium via the proton–proton chain.

Viewed from the outside, those two types of stars are basically similar, differing only quantitatively in size, temperature, and luminosity. Inside, they are qualitatively quite different because the upper-main-sequence stars are more massive and thus must have much higher central temperatures to be able to stand against their own gravity. Those high central temperatures allow these stars to fuse hydrogen via the CNO cycle, and that causes a completely different internal structure than in lower-mass stars such as the Sun that fuse hydrogen via the proton–proton chain.

You have learned that the CNO cycle is extremely temperature sensitive. To illustrate, if the central temperature of the Sun rose by 10 percent, energy production by the proton–proton chain would rise by about 46 percent, but energy production by the CNO cycle would shoot up 350 percent. This means that massive stars generate almost all of their energy in a tiny region at their very centers where the temperature is highest. A 10-solar-mass star, for instance, generates 50 percent of its energy from its central 2 percent of mass.

This concentration of energy production at the very center of the star causes a "traffic jam" as the energy tries to flow away from the center. Transport of energy by radiation can't drain away the energy fast enough, and the central core of the star churns in convection as hot gas rises upward and cooler gas sinks downward. Farther from the center, the traffic jam is less severe, and the energy can flow outward as radiation. This means that massive stars have convective cores at their centers and radiative envelopes extending from their cores to their surfaces (**Figure 11-13**).

The cores of main-sequence stars with less than about 1.1 solar masses cannot get hot enough to fuse much hydrogen via the CNO cycle. They generate nearly all of their energy by the proton–proton chain, which is not as sensitive to temperature, and thus the energy generation occurs in a core region with a relatively large diameter. The Sun, for example, generates 50 percent of its energy in a region that contains 11 percent of its mass. Because the energy generation is not concentrated at the very center of the star, no radiation traffic jam develops, and the energy flows outward easily in the form of light. Only near the surface, where the gas is cooler and therefore more opaque, does a radiation traffic jam develop, convection stirs the material, and energy flows outward in the form of gas motion rather than light. Consequently, the less massive stars on the main sequence, including the Sun, have radiative cores and convective envelopes, which is the opposite of higher-mass stars.

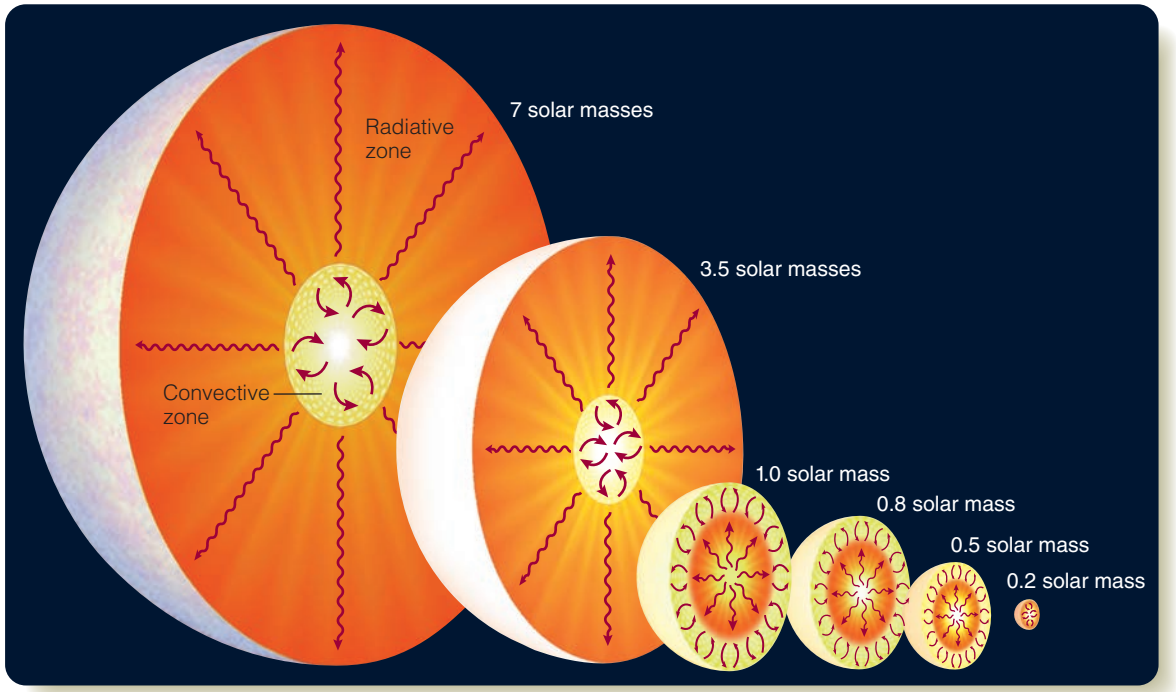
The lowest-mass stars have yet another kind of structure. For stars less than about 0.4 solar mass, the gas is relatively cool compared to gas inside more massive stars, so it has higher opacity and the radiation cannot flow outward easily. As a result, the entire bulk of these low-mass stars is stirred by convection.

The stars in the evening sky look much the same, but you have discovered that they are a diverse group, both outside and inside. Different kinds of stars make their energy in different ways and have different internal structures. Those structures are determined by the principle of balance. But how do stars maintain their stability? Can a star "lose its balance"?

11-5d The Pressure–Temperature Thermostat

Newborn stars contract and heat up until nuclear fusion begins. The energy flowing outward from the core heats the layers of gas, raises the pressure, and stops the contraction. This leads to an interesting question: How does the star manage to make just enough energy to stop contracting but not to start expanding? The key is the **pressure–temperature thermostat**, which is the relationship between pressure and temperature of gas that acts to keep the star burning steadily.

Consider what would happen if the reactions begin to produce too much energy. Normally, the nuclear reactions generate



▲**Figure 11-13** Inside stars. The more massive stars have small convective cores and radiative envelopes. Lower-mass stars, including the Sun, have radiative cores and convective envelopes. The lowest-mass stars are convective throughout. The "cores" of the stars where nuclear fusion occurs (*not shown*) are smaller portions of the interiors.

just enough energy to balance the inward pull of gravity. If the star makes slightly too much energy, the extra energy flowing out of the star would force its layers to expand slightly, lowering the central temperature and density and slowing the nuclear reactions until the star regained stability. Thus, a star has a built-in regulator that keeps its nuclear reactions from occurring too rapidly.

The same thermostat keeps the reactions from dying down. Suppose the nuclear reactions begin making too little energy. Then the star would contract slightly, increasing the central temperature and density, which would in turn increase the nuclear energy generation until the star regained stability.

The stability of a star depends on this relation between pressure and temperature. If an increase or decrease in temperature produces a corresponding change in pressure, the thermostat functions correctly, and the star is stable. You will discover in the next chapter how the pressure–temperature thermostat accounts for the relationship between mass and luminosity of main-sequence stars. In a later chapter, you will see what happens to a star when the thermostat breaks down, the star loses its internal balance, and the nuclear fires rage unregulated.

Practicing **SCIENCE**

What would happen if the Sun stopped generating energy? One of the ways scientists test their understanding is by doing thought-experiments, that is, imagining a situation and then considering the consequences of altering it somehow. If the thought-experiments are clever enough, they can substitute for real experiments on real stars, which of course are not possible.

Astronomers understand that stars are supported by the outward flow of energy generated by nuclear fusion in their interiors. That energy keeps each layer of the star just hot enough for the gas pressure to support the weight of the layers above. Each layer in the star must be in hydrostatic equilibrium; that is, the inward weight must be balanced by outward pressure. If the Sun stopped making energy in its interior, nothing would happen at first, but over many thousands of years the loss of energy from its surface would reduce the Sun's ability to withstand its own gravity, and it would begin to contract. You wouldn't notice much for 100,000 years or so, but gradually the Sun would lose its battle with gravity.

Stars are elegant in their simplicity—nothing more than a cloud of gas held together by gravity and warmed from the inside by nuclear fusion. Now imagine a different scenario: ***What would happen if the Sun's core suddenly increased its energy output?***

WHAT ARE WE?

Explainers

On cold winter nights when the sky is clear and the stars are bright, Jack Frost paints icy lacework across your windowpane. That's a fairy tale, of course, but it is a graceful evocation of the origin of frost. We humans are explainers, and one way to explain the world around us is to create myths.

An ancient Aztec myth tells the story of the origin of the Moon and stars. The stars, known as the Four Hundred Southerners, and the Moon-goddess Coyolxauhqui, plotted to murder their unborn brother, the great war god Huitzilopochtli. Hearing their plotting, he leaped from the womb fully armed, hacked Coyolxauhqui into pieces and chased the stars away. You can see the Four Hundred Southerners scattered across the sky, and each month you can see the Moon chopped into pieces as it passes through its phases.

Stories like these explain the origins of things and can make our Universe seem more understandable. Science is a natural extension of our need to explain the world. The stories have become sophisticated scientific hypotheses and theories that are tested over and over against reality, but we humans build those theories for the same reason people used to tell myths.

Study and Review

SUMMARY

11-1 Making Stars from the Interstellar Medium

- Stars are born from the gas and dust of the interstellar medium (ISM). Massive, hot stars like the prominent belt stars of Orion do not live very long lives compared to very low mass, cool stars. Their existence is strong evidence of star-forming regions.
- When enough mass has accumulated, gravity in giant molecular clouds makes the clouds contract. Gravity is resisted by thermal energy, magnetic fields, cloud rotation, and turbulence of the gas. If gravity wins, **dense cores** may form that are the first stage of star formation. In some cases, clouds are compressed by passing shock waves (also called shocks), triggering star formation. Shock waves can also be generated by the birth of massive stars, which trigger further star formation.
- In addition to high density, cores need high temperatures to initiate star formation. The cold gas of interstellar space heats up as it contracts because the atoms fall inward in **free-fall collapse** and pick up speed, converting gravitational energy into thermal energy. This thermal energy can leak out of the contracting cloud, allowing the rapid collapse to continue. Evidence of this stage is spectral emission **cooling lines** at wavelengths that can easily pass out of the contracting cloud.
- When a collapsing cloud becomes sufficiently dense and warm, far-infrared cooling lines are no longer effective and the cloud's contraction slows. A **protostar** forms inside this dusty **cocoon nebula** and is not directly detectable at visual wavelengths until the cocoon nebula dissipates.

11-2 The Orion Nebula: Evidence of Star Formation

- The visible Orion Nebula is only a small part of a much larger dusty molecular cloud that you cannot see. We see the Orion Nebula because ionization by ultraviolet photons from the hottest star lights up the nebula. **Bok globules** are small dark nebulae, some of which may be contracting to form stars.
- A single, very hot, short-lived O star in the Orion Nebula is almost entirely responsible for producing the ultraviolet photons that ionize the surrounding gas to make the nebula glow in visible light.
- Infrared observations reveal clear evidence of active star formation deeper in the molecular cloud just to the northwest of the Trapezium cluster, which is in the center of the Orion Nebula.
- Many young stars in the Orion Nebula are surrounded by disks, which are made of gas and dust. Such disks do not last long and are clear evidence that these stars are very young.
- **Star-formation pillars** are formed when ionized hot gas rushes away from newborn massive stars and encounters denser blobs of gas and dust.

11-3 Young Stellar Objects and Protostellar Disks

- Protostars become visible as their cocoon nebulae disappear and they cross the **birth line** in the H–R diagram. **T Tauri stars** have just emerged from their cocoon nebulae and are located in the H–R diagram between the birth line and the main sequence. Stars in the H–R diagram that are nearing the main-sequence stage are generally termed **Young Stellar Objects (YSOs)**.
- Many, perhaps most, protostars form surrounded by dusty **protostellar disks**, and jets of gas can be emitted as **bipolar flows** along the axis of the spinning disk. Where the jets push into the surrounding gas, they can form emission nebulae called **Herbig–Haro objects**. Protostellar disk material accretes into protostars, producing extra luminosity and adding to the stellar masses. Episodes of enhanced accretion flow and resulting luminosity outbursts are called the **FU Orionis phenomenon**.
- The presence of astronomically short-lived **associations** in a region, including **T associations** and **OB associations**, is evidence of recent star formation there. Protostars and YSOs produce **stellar winds** and **radiation pressure** that can blow their remaining disk and cocoon nebula material away.

11-4 Stellar Structure

- The law of **hydrostatic equilibrium** says that the weight of the gas above a location in a star must be balanced by the pressure at that location. Thus, inner layers of stars must be hotter because they must support more weight.
- The law of **energy transport** states that energy must flow from hot regions to cooler regions by conduction, radiation, or convection. Energy flows inside stars from the hot core in the center of the star to the cooler surface of the star by one of these mechanisms. In regions where the **opacity** of the gas is high and thus does not permit radiation to carry away enough energy, the gas can churn by convection.

11-5 The Source of Stellar Energy

- Many stars make their energy the same way the Sun does using the proton–proton chain. This fusion chain operates only at temperatures above 4 million K, the temperature needed to overcome the Coulomb barrier of electrical repulsion between positively charged atomic nuclei.
- The **CNO cycle** requires a higher temperature than the proton–proton chain because of the larger Coulomb repulsion by the nuclei in the CNO cycle. Both processes fuse four hydrogen nuclei to make one helium nucleus plus energy and subatomic particles.
- The CNO cycle is highly temperature sensitive, so stars more massive than 1.1 solar masses have centers hot enough to make energy via the CNO cycle. Less massive stars, including the Sun, make energy mostly via the proton–proton chain.
- The temperature sensitivity of the CNO cycle causes energy production in upper-main-sequence stars to occur in a very small region near the center, and the cores of these stars are convective zones. These stars transport energy by radiation in their outer layers. The proton–proton chain is not very sensitive to temperature, so the energy generation of lower-main-sequence stars, including the Sun, is more widely spread through the star's core, and the deep interior transports energy by radiation. The outer layers of these stars are convective. The lowest-mass main-sequence stars are so cool that their gases are relatively opaque, and they are convective throughout.
- The **pressure–temperature thermostat** is the relationship between pressure and temperature of a gas that tends to keep a star stable and fusing material steadily.

REVIEW QUESTIONS

1. What properties of a dense molecular cloud core must change to form a star?
2. What four factors cause a cloud of interstellar matter to resist contraction?
3. Explain three different ways a giant molecular cloud can be triggered to contract.
4. During a forming star's free-fall collapse stage, what is falling, why is it falling, and where is it falling toward?
5. During free-fall collapse, what keeps the collapsing cloud's internal temperature from increasing?
6. What happens to prevent a collapsing cloud from contracting to zero radius? That is, what happens to slow the ongoing collapse?
7. When the collapse of a cloud core stops, in what stage of star formation is that object?
8. Are cooling lines emission (bright) or absorption (dark) spectral lines?
9. What evidence indicates: (a) the existence of recent star formation? (b) the existence of protostars? (c) that the Orion region is actively forming stars?
10. How does a contracting protostar convert gravitational energy into thermal energy?
11. How does the geometry of bipolar flows and Herbig–Haro objects support the hypothesis that rotating disks surround protostars?
12. What is the source of the energy emitted by a protostar? By a T Tauri star? By a YSO?
13. What two forces drive the cocoon nebula away from a protostar?
14. Describe the three ways thermal energy can be transported.
15. Describe the principle of hydrostatic equilibrium as it relates to the internal structure of a star.
16. Describe how energy generated in the core of a 1-solar-mass star gets to the star's surface.
17. How does the CNO cycle differ from the proton–proton chain? How is it similar?
18. Why is CNO a cycle, whereas the proton–proton chain is a chain?
19. How does the extreme temperature sensitivity of the CNO cycle affect the structure of stars?
20. How does energy transport differ in the interior of a high-mass star from that in a medium-mass star such as the Sun?
21. How does the pressure–temperature thermostat control the nuclear reactions inside stars?
22. If fusion slowed in the core of a main-sequence star, what would happen? Would the star be in hydrostatic equilibrium?
23. **How Do We Know?** How would you respond to someone's comment about star formation, "That's only a theory"?
24. **How Do We Know?** Why can't scientists prove a scientific theory is totally correct?

ACTIVE INQUIRY QUESTIONS

1. Ancient astronomers, philosophers, and poets assumed that the stars were eternal and unchanging. Is there any observation they could have made or any line of reasoning that could have led them to conclude that stars don't live forever?
2. How does hydrostatic equilibrium relate to hot-air ballooning?
3. What is your favorite home-cooked meal? In terms of the three types of energy transport, how was it cooked?

PROBLEMS

1. A typical dense core has a diameter of 0.1 pc and a mass of 1 solar mass. What is the average density of this cloud core? Could this cloud core theoretically float on water? Does this object live up to its name—a *dense* core? (*Hint*: Assume the cloud core is spherical.) (*Notes*: The volume of a sphere is $\frac{4}{3}\pi r^3$; the density of water is 1000 kg/m^3; 1 pc $= 3.1 \times 10^{16}$ m. The mass of the Sun can be found in the Chapter 8 Celestial Profile: The Sun.)
2. The expanding bubble of hot gas shown in Figure 11-3b, which has been inflated by the light from a cluster of new stars in the bubble's center, has a diameter of about 100 ly. If the bubble is 170,000 ly from Earth, what is the observed diameter of the bubble in arc seconds? (*Hint*: Use the small-angle formula, Eq. 3-1.)
3. If a giant molecular cloud is 50 pc in diameter and a shock wave can sweep through it in 2 million years, how fast is the shock wave going in units of kilometers per second? (*Notes*: 1 pc $= 3.1 \times 10^{13}$ km; 1 yr $= 3.2 \times 10^7$ s.)

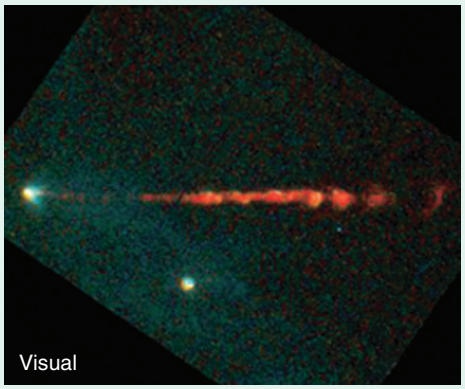
4. If a giant molecular cloud has a mass of 1.0×10^{35} kg and it converts 1 percent of its mass into stars during a single encounter with a shock wave, how many stars can it make? Assume the stars each contain 1 solar mass. (*Note*: The mass of the Sun can be found in the Chapter 8 Celestial Profile: The Sun.)
5. If a protostellar disk is 200 AU in radius and the disk plus the forming star together contain 2 solar masses, what is the orbital speed at the outer edge of the disk in kilometers per second? (*Hint*: Use the formula for circular velocity, Eq. 5-1a. Remember that the formula requires units of kg, m, and s.) (*Notes:* 1 AU $= 1.5 \times 10^{11}$ m; the mass of the Sun can be found in the Chapter 8 Celestial Profile: The Sun.)
6. During a free-fall collapse, the molecular cloud contracts, fragmenting into pieces. These fragments collapse further at about the same temperature as at the start of the molecular cloud's collapse, which is $T \sim 150$ K. Find the wavelength in nm at which the cloud will emit blackbody radiation most intensely. In which band of the electromagnetic spectrum is this radiation? (*Hint:* Use Wien's law, Eq. 7-2, and refer to Figure 6-3.)
7. If a contracting protostar is five times the radius of the Sun and has a temperature of only 2000 K, how luminous will it be relative to the Sun? (*Hint*: Use the luminosity-radius-temperature relation, Eq. 9-3.)
8. If a T Tauri star is the same temperature as the Sun but is ten times more luminous, what is its radius relative to the Sun? (*Hint*: Use the luminosity-radius-temperature relation, Eq. 9-3.)
9. The gas in a bipolar flow can travel as fast as 100 km/s. If the length of a jet is 1 ly, how long does a blob of gas take to travel from the protostar to the end of the jet? (*Notes*: 1 ly $= 9.5 \times 10^{12}$ km; 1 yr $= 3.2 \times 10^7$ s.)
10. Calculate the minimum surface temperature required for the single star in the Orion Nebula that is hot enough to ionize the hydrogen gas. According to this minimum surface temperature, what is the spectral class of this star? Does this spectral type agree with the information provided in the text about this star? (*Hint:* Use Wien's law, Eq. 7-2.) (*Notes:* The maximum wavelength of a photon that can ionize hydrogen is 91.2 nm. Appendix Table A-7 gives temperature versus spectral type.)
11. Circle all of the ^{1}H and ^{4}He nuclei in Figure 11-12 and explain how the CNO cycle can be summarized by $4\ {}^1\text{H} \rightarrow {}^4\text{He} + \text{energy}$.
12. Both the CNO cycle and the proton–proton chain combine 4 H nuclei to produce 1 He nucleus. Would those two processes release the same amount of energy per He nucleus produced? Why or why not? (*Hint:* Refer to Section 8-3.)
13. If the Orion Nebula is 8 pc in diameter and has a density of about 6.0×10^8 hydrogen atoms/m^3, what is its total mass? (*Notes*: The volume of a sphere is $\frac{4}{3}\pi r^3$; 1 pc $= 3.1 \times 10^{16}$ m; the mass of a hydrogen atom is 1.7×10^{-27} kg.)
14. If the hottest star in the Orion Nebula has a surface temperature of 40,000 K, at what wavelength does that star radiate the most energy? How does that compare with 91.2 nm, the wavelength of photons with just enough energy to ionize hydrogen? (*Hint*: Use Wien's law, Eq. 7-2.)

SENSE OF PROPORTION

1. The diameter of a typical interstellar cloud core with about the mass of the Sun is about 0.1 parsec. The diameter of the Sun is about 0.01 AU. How much larger than the Sun was the original cloud from which it formed? (*Note*: Conversions from parsecs to other units can be found in Appendix Table A-6.)
2. The age of the Sun is 4.6 billion years; the Sun's total main-sequence lifetime is calculated to be 11.0 billion years. In contrast, the time required for the Sun to form, from the start of its birth cloud's contraction until it became a hydrogen-burning main-sequence star, is calculated to be 30 million years. If the lifetime of the Sun were represented by the length of a football field, 100 yards = 300 feet, what length represents the Sun's current age? What length represents the time for the Sun to form?

LEARNING TO LOOK

1. Identify in Figure 11-4b the cool protostars, the cocoon nebulae, the giant molecular cloud, and general ISM material by color and location.
2. Using Figures 11-4a and 11-7, determine the spectral types, luminosities, colors, and surface temperatures of the forming Sun when passing through the protostar, T Tauri, and YSO stages.
3. In Figure 11-6, a dark globule of dusty gas is located at top right. What do you think that globule would look like if you could see it from the other side?
4. Compare the image of the Orion Nebula in Figure 11-4b with that in Figure 1 of the Concept Art: Star Formation in the Orion Nebula. Why are these two images different?
5. Locate the star-formation pillars in Figure 11-5. What are they pointing to?
6. The star shown in this photo appears to be ejecting a jet of gas. What is happening to this star?



NASA/ESA/STScI/AURA/NSF

12 Stellar Evolution

NASA/ESA/STScI/AURA/NSF/J.Hester(ASU)

▲ Eta Carinae is actually two massive post-main-sequence stars in a binary system. Their powerful luminosity is driving strong winds and mass loss, producing a disk plus two lobes of ejected material like a plate pressed between two basketballs. The larger component of Eta Carinae is expected to explode as a supernova within a few million years.

You learned in the previous chapter how stars form by condensing from dense clouds in the interstellar medium, then reach stability by fusing hydrogen into helium in their cores, releasing enough energy to counteract gravity. This chapter is about the subsequent long, stable middle age of stars on the main sequence and their old age as they swell to become giant stars. Here you will find answers to four important questions:

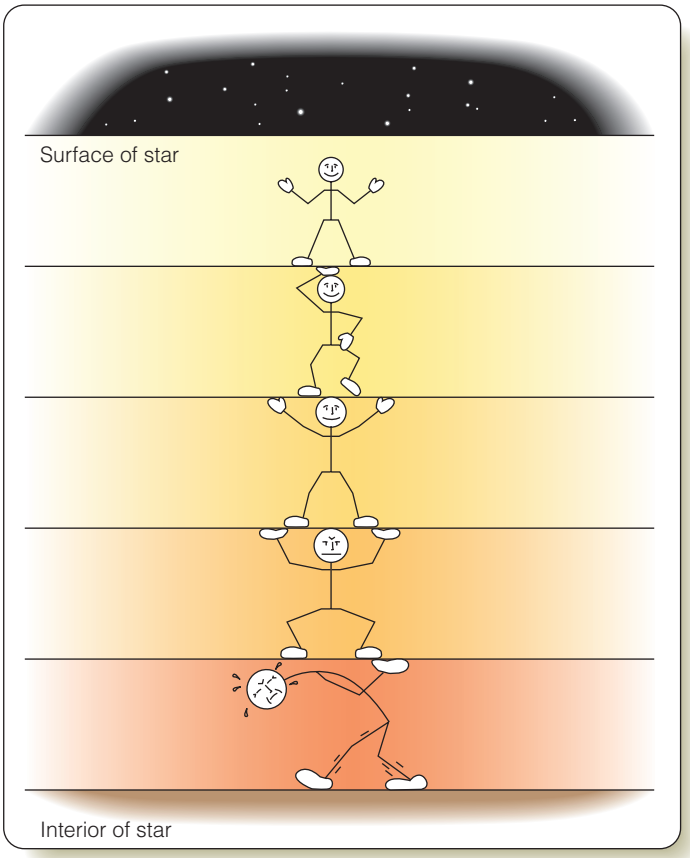
- Why is there a main sequence of star properties?
- Why is there a relationship between masses and luminosities of main-sequence stars?
- How does a star's structure change as it uses up its hydrogen fuel?
- What is the evidence that stars actually evolve?

This chapter is about how stars live. The next two chapters are about how stars die and the strange corpses they leave behind.

> We should be unwise to trust scientific inference very far when it becomes divorced from opportunity for observational test.
>
> —SIR ARTHUR EDDINGTON, *THE INTERNAL CONSTITUTION OF THE STARS*

EVERY STAR HAD A BEGINNING, and every star will someday have an ending. Between those events, the stars produce most of the light and energy that make our Universe beautiful. The stars above you seem eternal, but their existence depends on the fusion of nuclear fuels in their cores. Even as you watch, they are using up their fuel and drawing closer to their ends.

In this chapter, you will see the scientific method's full interplay of hypothesis and evidence used to understand how stars change as they consume their nuclear fuels. As the quotation that opens this chapter from one of the 20th century's greatest astrophysical theorists warns, being able to make clever hypotheses is not enough. At each step, astronomers, like all scientists, must confront hypotheses with evidence.



▲ **Figure 12-1** *Structure* in a star refers to the temperature, density, pressure, and so on, in each layer. Because each layer, like an acrobat in a circus stunt, must support the weight of everything above, astronomers can compute the conditions in each layer from the surface down to the center. Compare with Figure 11-10.

12-1 Main-Sequence Stars

The insides of stars are pretty much inaccessible to direct observation. Astronomers are limited to testing their hypotheses about stellar interiors against reality by making observations of only the exteriors of stars and testing hypotheses about stellar evolution with the "snapshots" of stellar lives available during the few centuries of modern astronomy. Despite those disadvantages, one of the greatest triumphs of modern astronomy has been the understanding, by mere human beings, that stars are not eternal but come into existence, evolve, and pass away.

12-1a Stellar Models

Astronomers can build mathematical models that describe the insides of stars, and the key to those models is balance. As you learned in Section 11-4a, a stable star is in hydrostatic equilibrium, balanced throughout its interior between two opposing forces: gravity that tries to make it contract, and internal pressure that tries to make it expand. By defining interior layers (called "shells") mathematically, astronomers can calculate the conditions at different positions inside the star, referred to as the "structure" of the star (**Figure 12-1**).

The internal structure of a star can be described by four simple laws of physics, two of which you have already encountered (Sections 11-4a and b). The law of hydrostatic equilibrium describes the balance between weight and pressure, and the law of energy transport describes how energy flows from hot to cool regions by radiation, convection, or conduction.

To those two laws you can add two basic laws of nature. When applied to stars, the **law of mass conservation** says that the total mass of the star must equal the sum of the masses in its shells with the added requirement that the mass be smoothly distributed through the star. No gaps are allowed. The **law of energy conservation** says that the amount of energy flowing out from the top of a shell in the star must be equal to the amount of energy coming in at the bottom of the shell plus whatever energy is generated within the shell. This means that the energy leaving the surface of the star—its luminosity—must equal the sum of the energy generated in all of the shells inside the star. This is like saying that the total number of new cars driving out of a factory must equal the sum of cars manufactured on all the assembly lines inside. No car can vanish into nothing or appear from nothing. Energy in a star may not vanish without a trace or appear out of nowhere.

Table 12-1 The Four Laws of Stellar Structure

1. Hydrostatic equilibrium	The weight on each layer is balanced by the pressure in that layer.
2. Energy transport	Energy moves from hot to cool regions by radiation, convection, or conduction.
3. Mass conservation	Total mass equals the sum of masses in all the layers. No gaps are allowed.
4. Energy conservation	Total luminosity equals the sum of energy generated per second in all the layers.

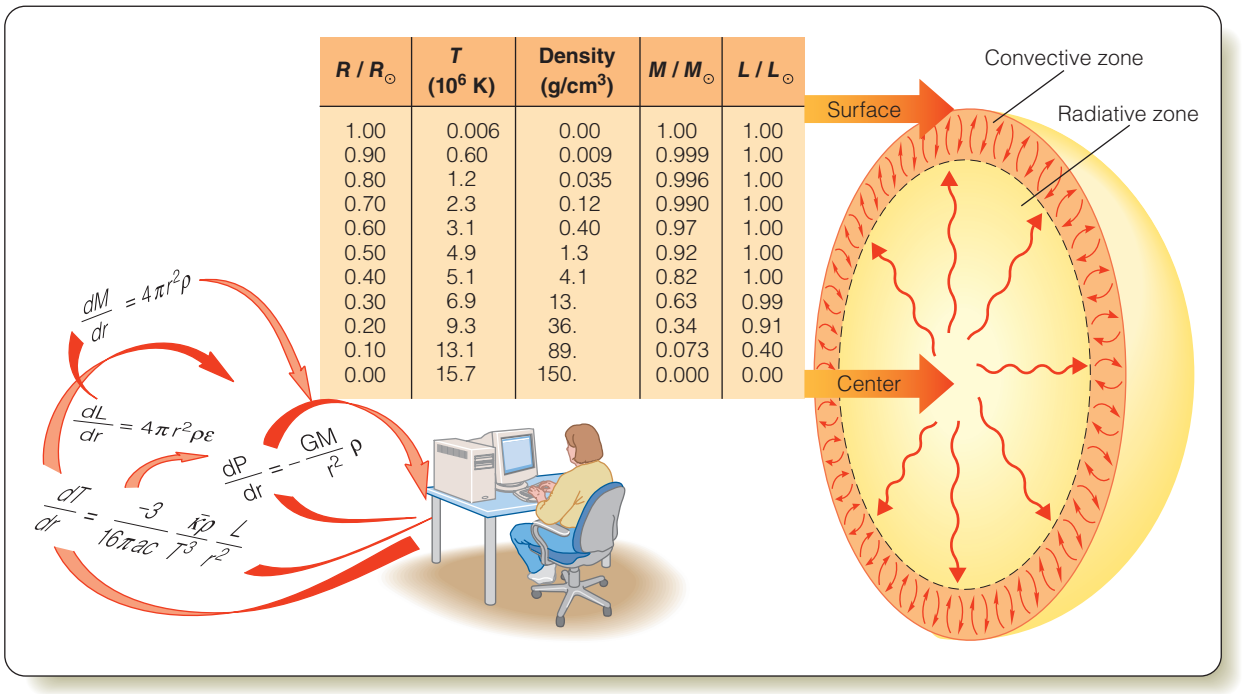
The four laws of stellar structure, described in qualitative terms in **Table 12-1**, can be written as mathematical equations. By solving those equations numerically using computer programs, astronomers are able to build mathematical models of the insides of stars.

If you want to calculate an accurate model of a star's interior, you have to divide the star's volume into at least 100 concentric shells and then write down the four equations of stellar structure for each shell. You would then have 400 equations that would have a total of 400 unknowns, namely, the temperature, density, mass, and energy flow in each shell. Solving 400 equations simultaneously is not easy, and the first such solutions, calculated by hand before the invention of electronic computers, took months of work. Now a properly programmed computer can solve the equations for a simple model in a few seconds and print a table of numbers that represent the conditions in each shell of the star. Such a table is a **stellar model**.

The table shown in **Figure 12-2** is a model of the Sun containing only ten layers, but it is extracted from a model with many more layers. The bottom line in the table, for radius equal to 0.00, represents the center of the Sun, and the top line, for radius equal to 1.00, represents the surface. The other lines in the table show the temperature and density in each shell, the mass inside each shell, and the fraction of the Sun's luminosity that is flowing outward through the shell. You can use such a model to understand many things about the Sun. For example, the model shows that all of the Sun's energy is generated near the center. No energy is generated in the outer layers. As you learned in Sections 8-1f and 8-3d, modern techniques of helioseismology and neutrino detection allow direct observational checks and fine-tuning of solar interior models, opportunities that would have amazed astronomers who constructed the first solar and stellar models in the early 20th century.

Notice that stellar models are quantitative; that is, properties have specific numerical values. Previously in this book, you studied models that were qualitative—descriptive, but not quantitative—for example, the Babcock model of the Sun's magnetic cycle. Both kinds of models are useful, but a quantitative model can reveal deeper insights into how nature works because it incorporates the precision of mathematics (**How Do We Know? 12-1**).

Stellar models let astronomers understand the inside of a star, and they can also look into a star's past and future. In fact, astronomers can use models like a time machine to follow the evolution of a star over billions of years. To look into a star's



▲ **Figure 12-2** A stellar model is a table of numbers that represent conditions inside a star. Such tables can be computed using the four laws of stellar structure, shown here in mathematical form. The table in this figure describes the present-day Sun.

HOW DO WE KNOW? 12-1

Mathematical Models

How can scientists study aspects of nature that cannot be observed directly?

One of the most powerful methods in science is the mathematical model, a group of equations carefully designed to describe the behavior of objects and processes that scientists want to study. Astronomers build mathematical models of stars to study the structure hidden deep inside them. Models can allow you to imagine speeding up the slow evolution of stars or slowing down the rapid processes that generate energy. Stellar models are based on only four equations, but other models are much more complicated and may require many more equations.

For example, scientists and engineers designing a new airplane don't just build it, cross their fingers, and ask a test pilot to try it out. Long before any metal parts are made, mathematical models are created to test whether the wing design will generate enough lift, whether the fuselage can support the strain, and whether the rudder and ailerons can safely control the plane during takeoff, flight, and landing. Those mathematical models are put through all kinds of tests: Can a pilot fly with one engine shut down? Can the pilot recover from sudden turbulence? Can the pilot land in a crosswind? By the time the test pilot rolls the plane down the runway for the first time, the mathematical models have flown many thousands of miles.

Scientific models are only as good as the assumptions that go into them and must be compared with the real world at every opportunity. If you are an engineer designing a new airplane, you can test your mathematical models by making measurements in a wind tunnel. Models of stars are much harder to test against reality, but they do predict some observable things. Stellar models predict the existence of a main sequence, the mass–luminosity relation, the observed numbers of giant and supergiant stars, and the shapes of star cluster H–R diagrams. Without mathematical models, astronomers would know little about the lives of the stars, and flying new airplanes would be a very dangerous business.

The Boeing Company

Before any new airplane flies, engineers build mathematical models to test its stability.

future, astronomers can use a stellar model to determine how fast the star uses the fuel in each of its shells. As the fuel is consumed, the chemical composition of the gas changes, the opacity changes, and the amount of energy generated declines. By calculating the rates of these changes, astronomers can compute a new model showing what the star will be like a few million years in the future. They can then repeat the process over and over, following the evolution of the star step-by-step as it ages over billions of years.

Modeling stellar structure and stellar evolution are highly challenging problems involving nuclear and atomic physics, thermodynamics, and sophisticated computational methods. Only since the 1970s have computers made rapid calculation of stellar models possible, and many of the advances in astronomy since then have been heavily influenced by the use of such models. The summary of star formation in the previous chapter is based on thousands of stellar models. You will continue to rely on theoretical models as you study the lives of main-sequence stars in the next section and the deaths of stars in the next chapter.

12-1b Why Is There a Main Sequence?

Astronomers have confidence in their stellar models because the laws of physics that go into the models are well understood. Another reason for confidence is that the models are constantly checked against reality by comparing them with observations of real stars. With that confidence, astronomers can use the models to understand stars better. For example, the models explain why there is a main sequence.

Models of stars show that there is a main sequence because stars support their weight by generating energy in their interiors. Gravity pulls all of the atoms in a star inward, but their weight is balanced by the outward pressure of the hot gas inside the star. The gas is heated by nuclear reactions, and the outward flow of that energy keeps each layer hot enough to support the weight pressing down on it. As a star forms, the first nuclear fuel to fuse is deuterium, which is the heavy isotope of hydrogen, because it has the lowest "ignition" temperature. However, deuterium is rare and produces relatively little energy. Stellar models show that deuterium fusion has little effect on halting the contraction of a protostar. Ordinary hydrogen fusion is the big powerhouse, and when it begins, it produces enough energy to balance gravity and stop the star's contraction. The main sequence is the location in the H–R diagram for stars in equilibrium, fusing hydrogen into helium.

There is a related question about the main sequence that you can resolve by thinking about stellar models. Why does the luminosity of a main-sequence star depend on its mass? Previously, you learned how observations of pairs of stars orbiting each other

in binary systems can be used to find their masses and discovered that the masses of main-sequence stars are ordered along the main sequence (look back to Section 9-6b). The least massive stars are at the bottom of the main sequence, and the most massive stars are at the top. That means there is a direct relationship between the mass of a main-sequence star and its luminosity. The main-sequence mass–luminosity relation is one of the most fundamental observations in astronomy that leads to understanding of stellar interiors. Stellar models can tell you why the mass–luminosity relation must be the way it is.

The keys to understanding the mass–luminosity relation are the law of hydrostatic equilibrium, which says that pressure must balance weight, and the pressure–temperature thermostat that regulates energy production. You first encountered these two ideas in Sections 11-4a and 11-5d. Hydrostatic equilibrium means that more massive stars must have higher central pressures and temperatures because they have more weight pressing down on their inner layers. For example, model calculations indicate that, to be stable, a 15-solar-mass ($M_\odot$) star must have a central temperature of about 34 million K, more than twice that of the Sun.

Furthermore, because massive stars have hotter cores, their pressure–temperature thermostats are "set" higher. The nuclear fuel at the center of a 15-$M_\odot$ star fuses 15,000 times more rapidly than the fuel at the center of the Sun. The rapid reactions produce more energy, which flows outward toward the cooler surface, heating each level in the star and enabling it to support the weight pressing inward. When that energy reaches the surface, it radiates into space as the star's luminosity. The important point here is that a mass–luminosity relation exists because stars must support their weight by generating energy, and more massive stars have more weight to support.

The explanation for the main sequence is elegant in its simplicity, but you can go beyond that to discover more about the inner workings of stars. In astronomy as well as in other fields of science, studying extreme cases can tell you a lot about a phenomenon. To understand main-sequence stars even better, you can now examine the upper (high-mass) and lower (low-mass) ends of the sequence.

12-1c The Upper End of the Main Sequence: High-Mass Stars

Models of stellar structure give astronomers a way to think about the opposite ends of the main sequence, the most massive and least massive stars. Both are difficult to study; massive stars because they are rare, low mass-stars because they are faint.

Taken together, theory and observation predict that there is an upper limit to the mass of stars. Astronomers have not found any stars more massive than about 150 $M_\odot$, although a few stars are inferred to have formed with as much as 300 $M_\odot$ but lost most of their bulk quickly via strong stellar winds. Observations show that, as interstellar clouds contract, they tend to fragment and produce many stars. The more mass a gas cloud contains, the more likely it is to fragment. This means that there aren't many extremely massive stars because most large gas clouds break into smaller fragments and form multiple-star systems.

Even if an extremely massive star does begin to form, stellar models reveal that such massive stars are unstable. To support the tremendous weight in such stars, the internal gas must be very hot, and that means it must emit floods of radiation that flow outward through the star. The resulting pressure of the radiation blows gas away from the star's surface in powerful stellar winds (**Figure 12-3**). Models indicate, for example, that a 60-$M_\odot$ star sheds mass so rapidly that it can be reduced to less than 30 $M_\odot$ in only a million years.

That mass-loss process sets an upper limit on the masses of stars, but it is difficult to test the models because it is hard to find such massive stars. Most B and O main-sequence stars have masses in the range of 5 to 40 $M_\odot$. The census of stars presented in Section 9-6 indicates that stars at the upper end of the main sequence are very rare, so astronomers must search to great distances to find just a few. Nevertheless, some stars thought to be very massive have spectra containing blueshifted emission lines. Kirchhoff's laws tell you that emission lines come from excited low-density gas, and a Doppler blueshift means that gas must be coming toward Earth. In other words, those stars are losing mass rapidly.

The image that opens this chapter shows Eta Carinae, a much-studied massive star nearing the end of its life. Observations and models suggest it is actually a binary containing two massive stars of about 110 and 45 $M_\odot$. The larger star may have formed with about 150 $M_\odot$, but it is losing mass rapidly. An eruption in 1841 made Eta Carinae the second brightest star in the sky and ejected the two expanding lobes of dusty gas you can see in the

Star Shadows Remote Observatory and PROMPT/UNC (Steve Mazlin, Jack Harvey, Rick Gilbert, and Daniel Verschatse)

▲ **Figure 12-3** Nebula NGC 2359, popularly known as Thor's Helmet, is actually a bubble about 30 ly across blown into a molecular cloud by a fast wind from the bright, massive star near the bubble's center. The star producing the wind and exciting the nebula is an extremely hot giant, of a type known as a Wolf-Rayet star, that is thought to be in a brief pre-supernova evolutionary state. The blue-green color is from oxygen emission lines.

picture. Although the star has faded, it is still very active, and more recent eruptions have ejected jets plus an equatorial disk of gas and dust. Massive stars like the components of Eta Carinae are clearly unstable.

12-1d The Lower End of the Main Sequence: Low-Mass Stars

The lower end of the main sequence is also difficult to study because, although these stars are common, they are dim and difficult to find even when they are only a few light-years away. If a red dwarf from the lower end of the main sequence replaced the Sun, it would be not much brighter than a full moon.

Stellar models predict that there are starlike objects even fainter than red dwarfs. Objects less massive than 0.08 $M_\odot$ cannot get hot enough in their cores to ignite hydrogen fusion. These **brown dwarfs** can't ignite hydrogen, so they contract slowly, convert their gravitational energy into thermal energy, and radiate it away. (You first encountered brown dwarfs in Section 9-3b in the context of stellar spectral types.) A low-mass red dwarf containing 0.08 $M_\odot$ and fusing hydrogen has a surface temperature of about 2500 K, but less massive brown dwarfs should have even lower temperatures, giving them a dull muddy-red color—thus the term *brown dwarf*.

Brown dwarfs are very difficult to find because they are even fainter than low-mass main-sequence stars, but large surveys and infrared studies have turned up hundreds of them. Some are located in binary systems with normal stars (Figure 12-4a), but large numbers are free-floating objects without stellar companions. A binary brown dwarf system has been discovered that is only 2 parsecs away, making it the third-closest "star" system; this indicates brown dwarfs might be nearly as common as normal stars.

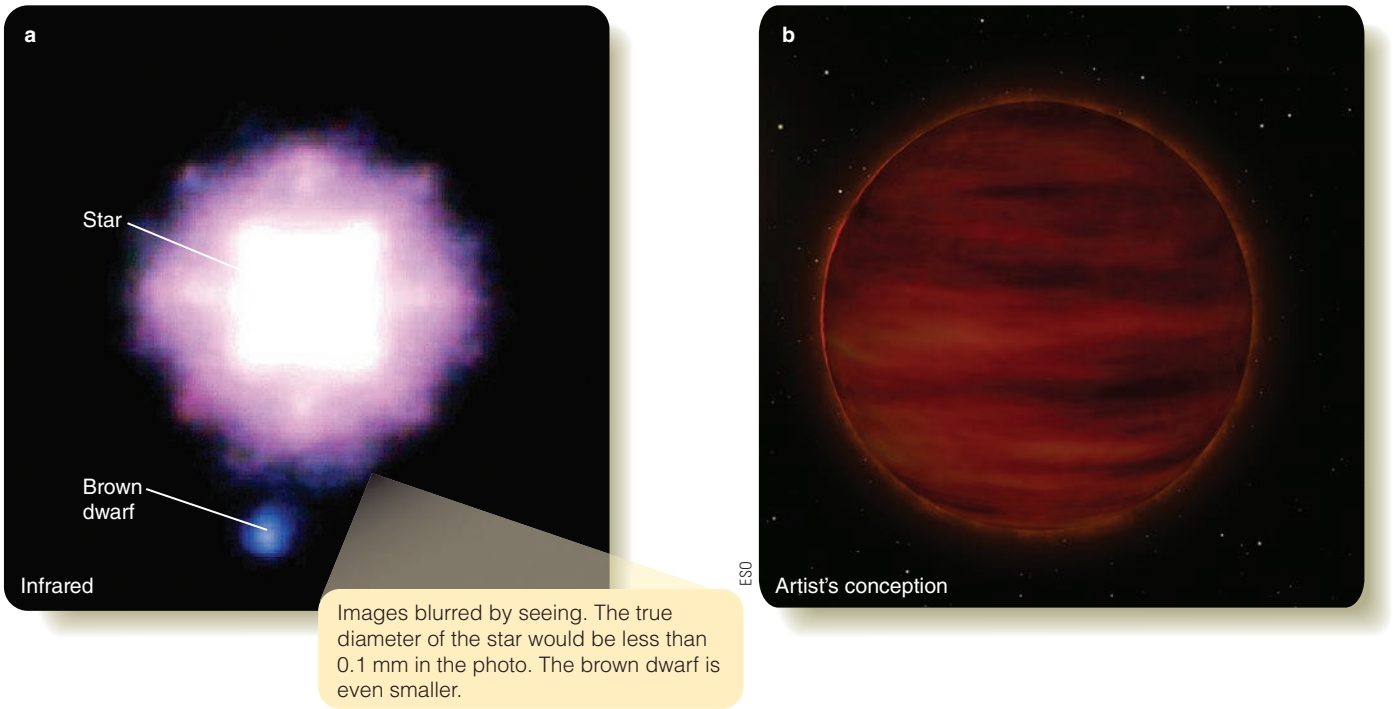
The L spectral type appears to overlap the coolest stars and the warmest brown dwarfs (look back to Figure 9-8). The T dwarfs are so cool they have methane bands in their spectra, like a giant planet. In fact, some brown dwarfs are so cool they can form solid grains and clouds in their atmospheres. Astronomers have observed color and brightness variations over time, which suggests that these brown dwarfs even have changing weather patterns (Figure 12-4b).

The lowest-mass brown dwarfs found so far are only a few times more massive than the planet Jupiter. It seems that brown dwarfs and giant planets like Jupiter are similar kinds of objects, although they probably formed by different processes. You will learn more about this in a later chapter.

12-1e The Life of a Main-Sequence Star

A normal main-sequence star supports its weight by fusing hydrogen into helium, but its supply of hydrogen is limited. As it consumes hydrogen, the chemical composition in its core changes, and the star evolves. Mathematical models of stars allow astronomers to follow that evolution.

As you know, hydrogen fusion combines four nuclei into one. Consequently, as a main-sequence star consumes its hydrogen, the total number of nuclei in its interior decreases.


a
Star
Brown dwarf
Infrared
Images blurred by seeing. The true diameter of the star would be less than 0.1 mm in the photo. The brown dwarf is even smaller.
b
Artist's conception

B.Biller and L.Close (Univ. of Arizona)

ESO

▲ **Figure 12-4** (a) Only 12.7 ly from the Sun, a brown dwarf orbits a low-mass M main-sequence star. Photographic effects give the brown dwarf a blue cast in the image, but if you could visit it, you would see an object slightly larger than the planet Jupiter glowing muddy red with a temperature slightly more than 1000 K. (b) Observations of brown dwarfs suggest that some have shifting weather patterns, as shown in the artist's conception.

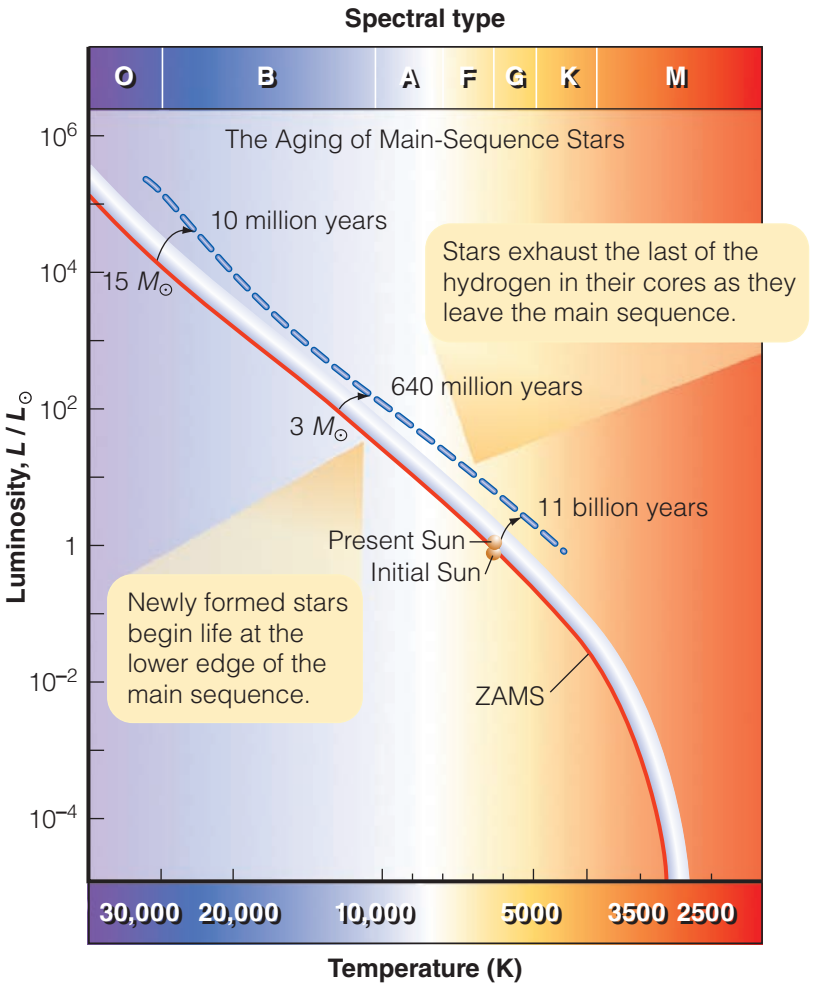
Each newly made helium nucleus exerts the same pressure as one hydrogen nucleus, but because the gas has fewer nuclei its total pressure is less. This unbalances the gravity–pressure stability, and gravity squeezes the core of the star more tightly. As the core slowly contracts, its temperature and density increase, and the nuclear reactions burn faster, releasing more energy. This additional energy flowing outward through the envelope forces the outer layers to expand and cool, making the star slightly larger, brighter, and cooler.

As a result of these gradual changes in main-sequence stars as they age, the main sequence is not actually a sharp line across the H–R diagram but rather a band (Figure 12-5). When a star begins its stable life fusing hydrogen, its properties put it on the lower edge of this band, called the **zero-age main sequence (ZAMS)**. As a star combines hydrogen nuclei to make helium nuclei, the point that represents the star's luminosity and surface temperature moves slightly upward and to the right, eventually reaching the upper edge of the main-sequence band just as the star exhausts nearly all of the hydrogen in its center. The important point here is that astronomers observe stars with properties evenly distributed across the main-sequence band, which is clear evidence verifying models that predict these slow changes as stars get older.

These gradual changes in the Sun will spell trouble for Earth. When the Sun began its main-sequence life almost 5 billion years ago, it had only about 70 percent of its present luminosity. By the time the Sun leaves the main sequence in another 6 billion years, it will have twice its present luminosity. As a result, the average temperature on Earth will have climbed by at least 50°C (90°F). As this happens over the next few billion years, first the polar caps will melt completely; then, eventually, the oceans will evaporate. (This is not related to the current global warming, which affects Earth's climate much more quickly and is not caused by the Sun.) Clearly, the future of Earth as the host of a biosphere is limited by the evolution of the Sun.

Once a star leaves the main sequence, it evolves rapidly and soon dies. The average star spends about 90 percent of its life on the main sequence, fusing hydrogen to helium. This explains why about 90 percent of all normal stars are main-sequence stars. You are most likely to see a star during its long, stable time on the main sequence.

How long a star spends on the main sequence depends on its mass. Massive stars have lots of fuel, but they use it rapidly and live short lives. Low-mass stars conserve their fuel and shine for billions of years (Table 12-2). For example, a 25-$M_\odot$ star will exhaust its hydrogen and die in only about 4 million years, whereas the Sun can fuse hydrogen in its core for about 11 billion years. This means, for one thing, that astronomers think



▲ **Figure 12-5** Contracting protostars reach stability with properties at the lower edge of the main sequence, called the zero-age main sequence (ZAMS). As a star converts hydrogen in its core into helium, its position in the H–R diagram moves slowly across the main sequence, becoming slightly more luminous and slightly cooler. Once a star consumes all of the hydrogen in its core, it can no longer remain a stable main-sequence star. More massive stars age rapidly, but less massive stars use up the hydrogen in their cores more slowly and live longer main-sequence lives. (Evolutionary tracks adapted from the work of Icko Iben.)

Table 12-2 Main-Sequence Stars

Spectral Type	Mass (Sun = 1)	Luminosity (Sun = 1)	Years on Main Sequence
O5	40	400,000	1×10^6
B0	15	13,000	11×10^6
A0	3.5	80	440×10^6
F0	1.7	6.4	3×10^9
G0	1.1	1.4	8×10^9
K0	0.8	0.46	17×10^9
M0	0.5	0.08	56×10^9

that life is unlikely to develop on planets orbiting massive stars. These stars do not live long enough for life to get started and evolve into complex creatures. You will learn more about this in the book's final chapter on life in the Universe.

Red dwarfs have low masses and use their fuel so slowly they should survive for hundreds of billions of years. As you will learn in a later chapter, the Universe is only about 14 billion years old, so red dwarf stars must all still be in their infancy. None of them have exhausted their hydrogen fuel yet.

Vast numbers of faint, low-mass stars fill the sky. Look back to the stellar census in the Chapter 9 Concept Art: The Family of Stars, and recall how much more common the lower-mass main-sequence stars are than the massive O and B stars. Main-sequence K and M stars are so faint they are difficult to locate, but they are very common. Studies of star formation show that nature makes more low-mass stars than massive stars, but that alone is not sufficient to explain the excess of low-mass stars. An additional factor is the stellar lifetimes. Because low-mass stars not only form more often but also live long lives, there are many more of them in the sky than massive stars. O and B main-sequence stars are luminous and easy to locate, but because they form rarely and have fleeting lives, there are never more than a few of them nearby at any one time.

12-1f The Life Expectancies of Stars

To really understand how stars evolve, you need to know how long they can survive. In fact, you can easily estimate the life expectancy of a star from its mass.

Because main-sequence stars consume their fuel at an approximately constant rate, you can estimate the amount of time a star spends on the main sequence—its life expectancy, symbolized by $t_\star$—by dividing the amount of fuel by the rate of fuel consumption. You have made calculations like this before. If you drive a truck that carries 30 gallons of fuel and uses 3 gallons of fuel per hour, you know the truck can run for 10 hours.

The amount of fuel a star has can be assumed to be proportional to its mass, and the rate at which it burns its fuel is proportional to its luminosity. This means that you can make a first estimate of the star's life expectancy by comparing its mass to its luminosity. For example, astronomers observe that 2-$M_\odot$ main-sequence stars are about 11 times more luminous than the Sun. If stars with 2 $M_\odot$ have twice as much fuel as the Sun but consume that fuel 11 times more quickly than the Sun does, they should live only about 2/11, or 18 percent, as long as the Sun. Notice that stars larger than the Sun actually have a shorter lifetime because, although they have more fuel, they consume that fuel much faster.

You can make the calculation even easier if you remember the mathematical form of the mass–luminosity relation (Eq. 9-5), which is that the luminosity of a main-sequence star is approximately proportional to $M^{3.5}$. The life expectancy of a main-sequence star then is:

KEY EQUATION 12-1 Stellar Life Expectancies

$$t_\star = \frac{M_\star}{L_\star} = \frac{M_\star}{M_\star^{3.5}} = \frac{1}{M_\star^{2.5}}$$

This means that you can estimate the life expectancy of a star, t_*, relative to the lifetime of the Sun ($t_\odot$) by taking the star's mass in solar units, raising it to the 2.5 power, and then taking the reciprocal of that. If you express the mass in units of solar masses, the life expectancy will be in units of solar lifetimes.

Example: How long can a 4-$M_\odot$ star live, relative to the Sun?

$$t_\star = \frac{1}{4^{2.5}} = \frac{1}{32}\, t_\odot$$

Solution: Detailed studies of models of the Sun show that the Sun, presently almost 5 billion years old, can last about another 6 billion years. So, a solar lifetime is approximately 11 billion years, and a 4-$M_\odot$ star will last about (11 billion)/32, or 340 million, years.

This way of estimating stellar life expectancies is very approximate. For example, the model ignores mass loss, which astronomers have evidence especially affects the life expectancies of the most massive and least massive stars. Nevertheless, it serves to illustrate an important point. Stars that are only somewhat more massive than the Sun have dramatically shorter lifetimes on the main sequence.

Practicing SCIENCE

Why is there a main sequence? This is the kind of question a scientist would ask after discovering that there is a simple relationship between mass and luminosity for main-sequence stars.

To be stable, the weight of a star's material must be balanced by pressure. This is the principle of hydrostatic equilibrium. Imagine a contracting protostar that is not quite in equilibrium, as gravity squeezes it tighter and tighter. As it contracts, its interior heats up, and when it gets hot enough to fuse hydrogen into helium, the pressure–temperature thermostat takes over and regulates energy production so that the star makes just enough energy to support its own weight.

Massive stars, having more weight to support, must have higher internal pressures and temperatures and therefore must be generating more energy and have a greater luminosity. Massive stars reach stability with luminosities and surface temperatures that place them along the upper main sequence on the H–R diagram. Less massive stars support less weight and can reach stability along the lower main sequence. There is a main sequence because stars in this stable life stage all fuse hydrogen to support their own weight.

Now consider a consequence of the mass–luminosity relation: ***Why do massive stars have such short life expectancies?***

12-2 Post-Main-Sequence Evolution

Why are giant stars so large, and why do they have such low densities? Why are they so uncommon? You are ready to answer those questions by learning about the evolution of stars after they leave the main sequence.

12-2a Expansion into a Giant

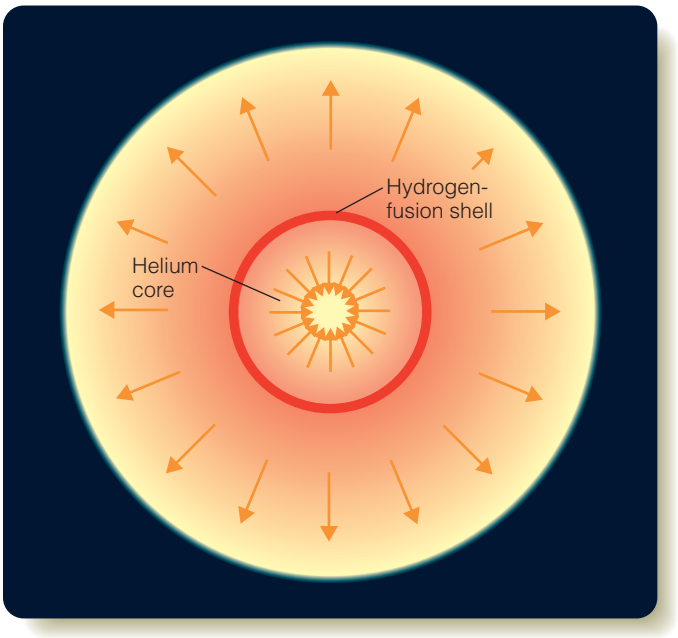
To understand how stars evolve, you might recall that the centers of medium-mass stars like the Sun are radiative, meaning the energy moves as radiation and not as circulating currents of heated gas. As a result, the gas does not move deep inside such stars, which means their interiors are not mixed at all. (The lowest-mass stars are an exception that you will examine in the next chapter.) More massive stars have convective cores that mix the central regions (Figure 11-13), but those regions are not very large, and so, for the most part, the interiors of massive stars also are not mixed.

In this respect, medium- and high-mass stars are like campfires that are not stirred: Ashes accumulate at the centers, and fuel in the outer parts never gets used. Hydrogen fusion produces helium nuclei—the "ashes" that accumulate at the star's center. Because nothing mixes the interior of the star, the helium nuclei remain where they are in the center, and the hydrogen in the star's outer parts is not sent down to the center where it could be fused.

The helium ashes that accumulate in a main-sequence star's core cannot fuse into heavier elements because the temperature is too low. As a result, the core eventually becomes an inert ball of helium. As this happens, the energy production in the core falls, and the weight of the outer layers forces the core to contract.

Although the contracting helium core cannot generate nuclear energy, it does grow hotter because it converts gravitational energy into thermal energy (Section 11-1b). The rising core temperature heats the unprocessed hydrogen surrounding the core—hydrogen that was never before hot enough to fuse. When the hydrogen just outside the core becomes hot enough, it ignites in a thin, spherical shell. Like a grass fire moving outward from an exhausted campfire, the hydrogen-fusing shell slowly burns outward, leaving more helium ash behind and increasing the mass of the helium core.

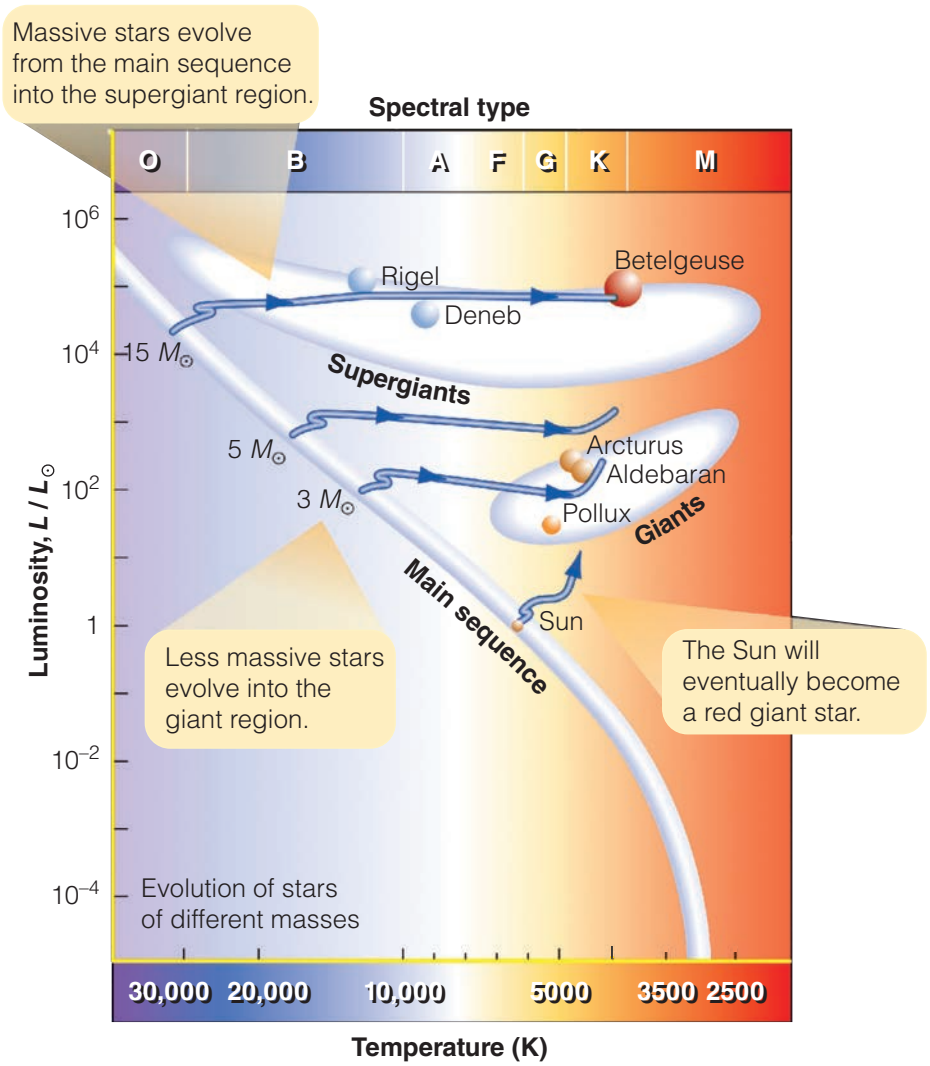
During this stage in its evolution, the star overproduces energy; that is, it produces more energy than it needs to balance its own gravity. The contracting helium core converts gravitational energy into thermal energy. Some of this energy heats the helium core, and some of that heat leaks outward through the star. At the same time, the hydrogen-fusion shell also produces energy as the contracting core brings unburned hydrogen closer to the center of the star and heats it to high temperature. The result is a flood of energy outward, forcing the outer layers of the star to puff up and expand the star into a giant (Figure 12-6). Note how, at this stage, the overall structure of the star is changing dramatically: The core is contracting while the outer envelope is expanding.



▲ **Figure 12-6** When a star runs out of hydrogen at its center, it ignites a hydrogen-fusion shell. The helium core contracts and heats while the envelope expands and cools. (For a scale drawing, see Figure 12-9.)

The expansion of the envelope changes the star's location in the H–R diagram. As the outer layers of gas expand, energy is absorbed in lifting and expanding the gas. The loss of that energy lowers the temperature of the gas. Consequently, the point that represents the star in the H–R diagram moves quickly to the right: in about 150 million years for a star of 1 $M_\odot$, but in less than a million years for a star of 5 $M_\odot$ (Figure 12-7). As the radius of the evolving star continues to increase, the enlarging surface area makes the star more luminous, moving its point upward in the H–R diagram. One of the Famous Stars, Aldebaran, the glowing red eye of Taurus the Bull, is a red giant, with a diameter more than 40 times that of the Sun but with only 2/3 of the Sun's surface temperature.

A medium-mass star like the Sun evolves to become a giant, but more massive stars evolve across the upper H–R diagram and become supergiants, which are stars even larger than giants. Consider two more Famous Stars. Betelgeuse (Alpha Orionis) is a very cool, red supergiant with about 1000 times the diameter of the Sun. Rigel (Beta Orionis) is a supergiant 80 times larger than the Sun. It may seem odd to say that Rigel, a blue star, has expanded and cooled. At a temperature of 12,000 K, Rigel is hot and looks quite blue to your eyes. (Any star as hot as, or hotter than, Rigel will look blue to your eyes, as shown by the



▲ **Figure 12-7** The evolution of a massive star moves the point that represents it in the H–R diagram to the right of the main sequence into the region of the supergiants such as Rigel and Betelgeuse. The evolution of a medium-mass star moves its point in the H–R diagram into the region of the giants such as those shown here. (Evolutionary tracks adapted from the work of Icko Iben.)

blackbody curves in Figure 7-7.) But when it was on the main sequence, Rigel had a much hotter surface. Even so, it would not have looked much bluer than it does now because you cannot see the ultraviolet radiation that the hottest stars emit. So, although Rigel is blue, it is in fact a star that has expanded and cooled.

Now you can understand a few things you noticed in previous chapters. Giants and supergiants are large because they have expanded, and they have very low densities for that reason. Also the giant and supergiant region of the H–R diagram contains a jumble of different mass stars (look back to Figure 9-23) because main-sequence stars end up in about the same part of the diagram as they age, no matter what their original masses were. Thus, there is no clear mass–luminosity relationship among the giant and supergiant stars.

12-2b Degenerate Matter

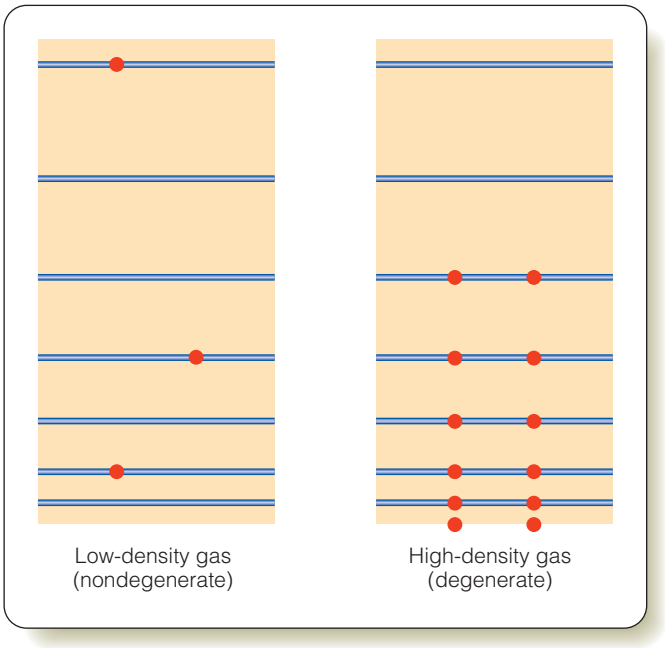
Although the hydrogen-fusion shell can force the envelope of the star to expand, it cannot stop the contraction of the helium core. Because the core is not hot enough to fuse helium, gravity squeezes it tighter, and it becomes very small. If you were to represent the helium core of a giant star with a baseball, the outer envelope of the star would be about the size of a baseball stadium, yet the relatively tiny core would contain roughly 12 percent of the star's mass. When gas is compressed to such extreme densities, it begins to behave in surprising ways that can strongly affect the evolution of a star.

Normally, the pressure in a gas depends on its temperature. The hotter a gas is, the faster its particles move, and the more pressure it exerts. The gas inside a star is ionized, so there are two kinds of particles, atomic nuclei and free electrons. Under normal conditions the gas in a star follows the same pressure–temperature laws as other gases, but if the gas is compressed to very high densities, as in the core of a giant star, two laws of quantum mechanics that determine the behavior of subatomic particles come into play, and the differences between electrons and nuclei becomes important.

First, quantum mechanics says that the moving electrons confined in the star's core can have only certain amounts of energy, just as the electron in an atom can occupy only certain energy levels (Section 7-1c). You can think of these permitted energies as the rungs of a ladder. An electron can occupy any rung but not the spaces between.

The second quantum mechanical law is called the Pauli exclusion principle, which says that two identical electrons cannot have the same location and occupy the same energy level. Because electrons spin in one direction or the other, two electrons can occupy a single energy level if they spin in opposite directions. That level is then completely filled, and a third electron cannot enter that level (cannot have that energy) because, whichever way it spins, it will be identical to one of the two electrons already in that level. The Pauli exclusion applies to electrons flying free in an ionized gas, not attached to any nucleus, as well as to electrons within an atom.

A low-density ionized gas has few electrons per cubic centimeter, so there are plenty of combinations of position and energy



▲ **Figure 12-8** Electron energy levels are arranged like rungs on a ladder. In a low-density gas, many levels are open, but in a degenerate gas all lower-energy levels are filled. That causes the strange behavior of degenerate matter.

level available (**Figure 12-8**). If a gas becomes very dense, however, nearly all of the lower energy levels in a given location are occupied. In such a gas, a moving electron cannot slow down; slowing down would decrease its energy, and there are no open lower energy levels into which it can drop. And it can speed up only if it absorbs a lot of energy, enough so it can move faster than *all* the other electrons and thereby immediately leap to the top of the energy ladder where there are empty energy levels. So, that electron is stuck; it can't speed up, and it can't slow down.

When a gas is so dense that most of its electrons are not free to change their energies, it is called **degenerate matter**. (Note that, to become degenerate, the gas must have a density on the order of 1 million g/cm^3. On Earth, a teaspoon of that material would weigh as much as a large truck.) Although it is a gas, it has two peculiar properties that can affect the star. First, the degenerate gas resists compression. To compress the gas, you must push against the moving electrons, changing their positions and motions. But changing their motions requires changing their energies. That requires tremendous effort because you can't just change their energies slightly; you must boost them completely to the top of the energy ladder where there are empty rungs. That is why degenerate matter, though still a gas, is harder to compress than the toughest hardened steel; if you push on it or squeeze it, nothing happens.

Second, the pressure of degenerate gas does not depend on temperature. To see why, note that the *pressure* of the gas depends mostly on the speed of the electrons because they are more numerous than the nuclei. But, as was just described, the speed of the electrons cannot be changed without tremendous effort. The *temperature* of the gas, however, depends on the motions of all the particles in the gas, both electrons and nuclei. If you add heat to the gas, most of that energy goes to speed up the motions of the nuclei, which are not degenerate, move relatively slowly, and don't contribute much to the pressure. Thus, raising the temperature of a degenerate gas only affects the motion of the nuclei and has almost no effect on the pressure.

These two properties of degenerate matter (it resists compression, and its pressure does not depend on the temperature) become important when stars end their main-sequence lives. Eventually, many stars collapse into tiny white dwarf remnants that are made of degenerate matter. But long before that, the cores of some giant stars become so dense that they are degenerate, a situation that can produce a cosmic bomb when helium begins to fuse.

12-2c Helium Fusion

Hydrogen fusion leaves behind helium ash, which cannot fuse at the relatively low temperatures of hydrogen fusion. Helium nuclei (two protons plus two neutrons) have a positive charge twice that of a hydrogen nucleus (a proton). Because of the greater electrical charge of helium nuclei than hydrogen nuclei, the Coulomb barrier to pushing helium nuclei together is higher than for hydrogen nuclei. At the temperature of hydrogen fusion, helium nuclei move too slowly and collide too gently to fuse.

As the star becomes a giant star, fusing hydrogen in a shell, its inner core of helium ash contracts and grows hotter. Finally, as the temperature approaches 100 million K, helium nuclei can begin to fuse together to make carbon. Because three helium nuclei are needed to make a carbon nucleus, and because a helium nucleus is also called an alpha particle, astronomers usually refer to helium fusion as the **triple alpha process**.

You can summarize the helium-fusing process in two steps:

$$^4\text{He} + {}^4\text{He} \rightarrow {}^8\text{Be} + \gamma$$

$$^8\text{Be} + {}^4\text{He} \rightarrow {}^{12}\text{C} + \gamma$$

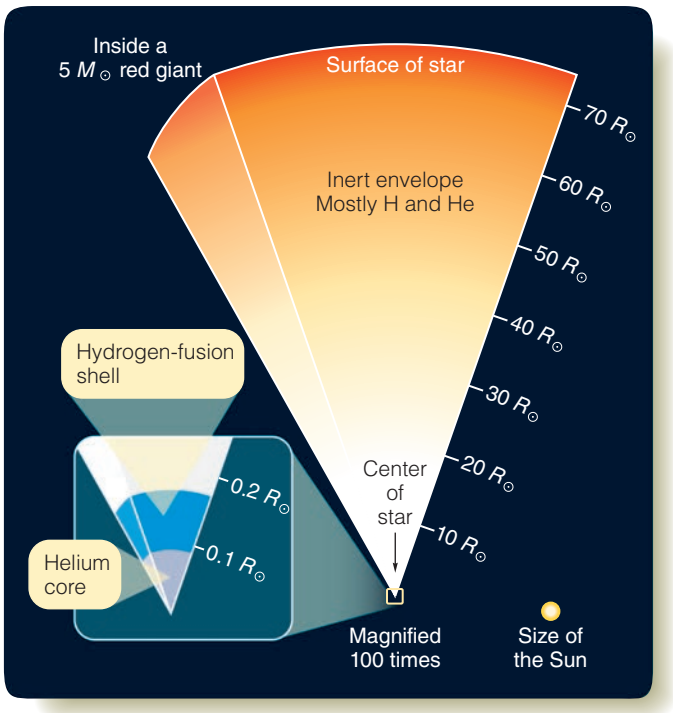
(The γ symbol represents a gamma-ray photon.) The first reaction actually absorbs a little bit of energy from the gas, but the second reaction produces almost 80 times more than the first reaction absorbs. This process is complicated by the fact that a beryllium-8 nucleus is very unstable and easily breaks apart back into two helium nuclei before it can absorb another helium nucleus. Three helium nuclei can also form carbon directly, but such a triple collision is extremely unlikely, so helium fusion depends on the more likely event of a third helium nucleus colliding with a beryllium-8 nucleus in the short interval before it breaks apart.

Some stars begin helium fusion gradually, but stars in a certain mass range begin helium fusion with an explosion called the **helium flash**. This explosion occurs if the gas in the helium core becomes degenerate. Because pressure no longer depends on temperature in a degenerate gas, the pressure–temperature thermostat that controls the nuclear fusion reaction rates no longer works.

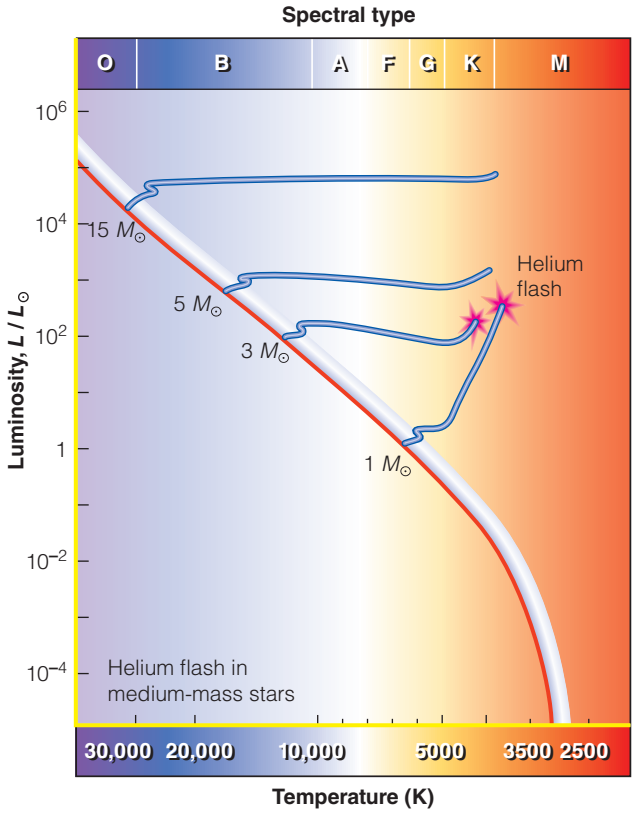
When the degenerate helium ignites, it generates energy, which raises the temperature. But because the pressure–temperature thermostat is not operating, the core does not respond to the higher temperature by expanding and reducing the temperature. Rather, the higher temperature forces the reactions to go faster, making more energy, which raises the temperature, which makes the reactions go faster, and so on. The ignition of helium fusion in the degenerate gas thus results in a runaway explosion so violent that the helium core temporarily generates more than 10^{11} times as much energy per second as does the Sun. That's comparable to the luminosity of all the stars in the Milky Way Galaxy.

Although the helium flash is sudden and powerful, it does not destroy the star. In fact, if you were observing a giant star as it experienced a helium flash, you would probably see no outward evidence of an eruption. The helium core is quite small (**Figure 12-9**), and much of the explosion's energy goes into heating the core or is absorbed by the star's expanded envelope. Also, the helium flash happens very quickly. In a matter of seconds the core of the star becomes so hot it is no longer degenerate, the pressure–temperature thermostat is able to bring the helium fusion under control, and the star proceeds to fuse helium slowly and steadily in its core.

Astronomers calculating models of stellar evolution find that the Sun and other medium-mass main-sequence stars will eventually undergo a helium flash, but not all stars do. Stars less massive



▲ **Figure 12-9** When a star runs out of hydrogen at its center, the core of helium contracts to a small size, becomes very hot, and begins nuclear fusion in a shell (blue). The outer layers of the star expand and cool. The red giant star shown here has an average density much lower than the air at Earth's surface. Here $M_\odot$ stands for the mass of the Sun, and $R_\odot$ stands for the radius of the Sun.



▲ **Figure 12-10** Stars like the Sun undergo a helium flash, but more massive stars begin helium fusion without a helium flash. Stars less than 0.4 times the mass of the Sun cannot get hot enough to ignite helium. The zero-age main sequence is shown here in red at the lower edge of the main sequence. (Evolutionary tracks adapted from the work of Icko Iben.)

than about 0.5 $M_\odot$ never get hot enough to ignite helium. Stars more massive than about 3 $M_\odot$ ignite helium before their cores become degenerate (**Figure 12-10**), so their pressure–temperature thermostat keeps the helium fusion under control and there is no helium flash.

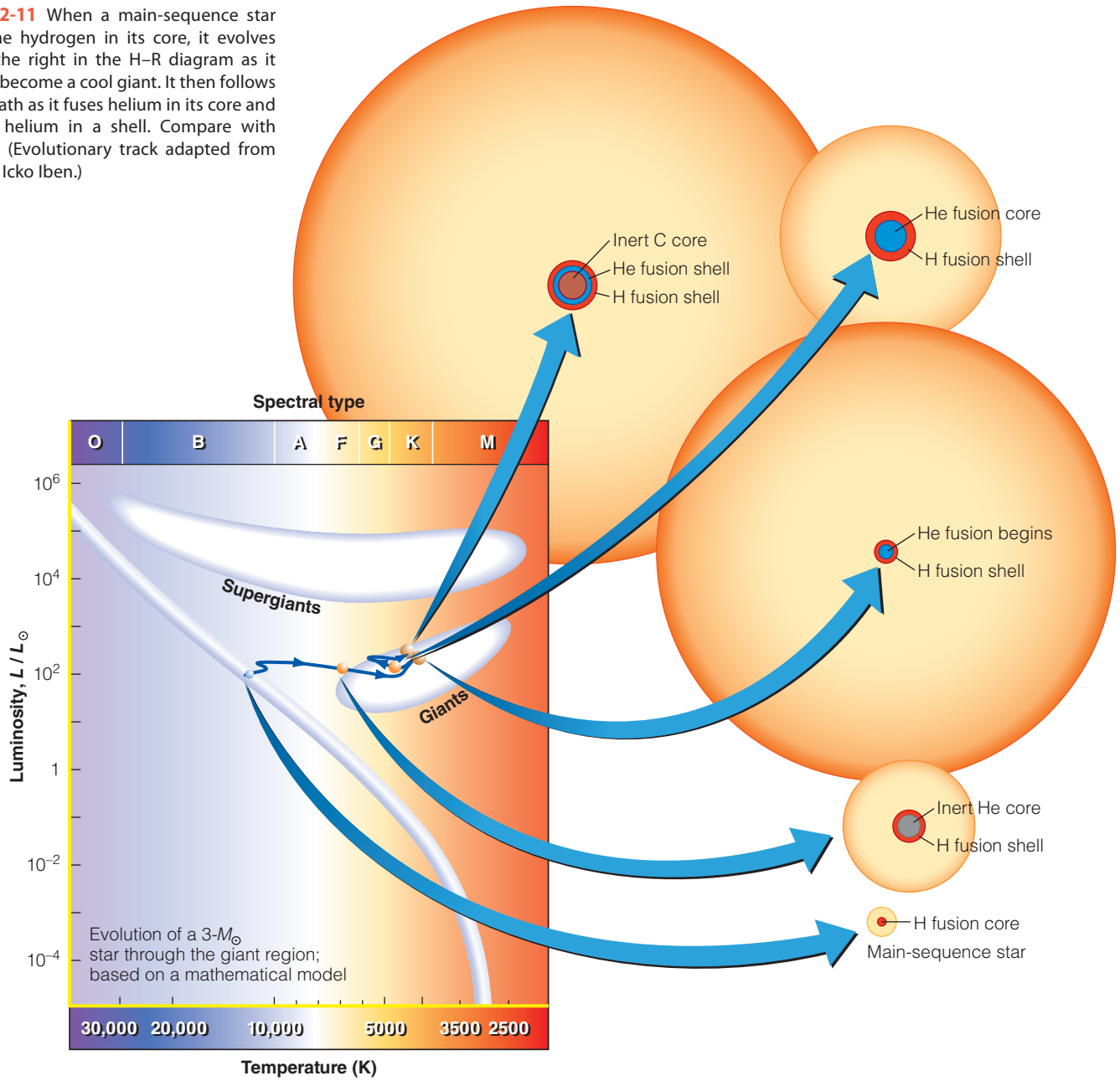
If the helium flash occurs only in some stars and is a very short-lived event that is invisible from outside the star, why is it important? The answer is that the helium flash phenomenon is an area of exceptional uncertainty in the study of stellar evolution for stars with about the mass of the Sun. Massive stars are not very common, and low-mass stars evolve slowly, so it is hard to gather observational evidence that can be used to check models of high-mass and low-mass stellar evolution. Studies of stellar evolution largely focus on medium-mass stars, and those are the stars that do experience the helium flash. But the helium flash occurs so rapidly and so violently that modeling programs cannot calculate the changes in a star's internal structure in sufficient detail. To follow the evolution of medium-mass stars like the Sun past the helium flash, astronomers must make guesses rather than exact calculations about the way the helium flash affects stellar structures.

There is another reason that the helium flash is important. Understanding what happens when the pressure–temperature

thermostat is turned off lets you appreciate its importance in maintaining the stability of stars. In the next chapter, you will meet stars that explode violently when their fusion reactions become completely unregulated.

On the other hand, post-helium-flash evolution is generally understood. After the flash, the stellar core expands, the gas ceases to be degenerate, and helium fusion proceeds under the control of the pressure–temperature thermostat. Also, expansion of the star's core absorbs energy that otherwise would support the outer layers of the star. As a result, the outer layers contract, and the surface of the star becomes somewhat hotter. In the H–R diagram, the point that represents the star initially moves down toward lower luminosity and then to the left toward higher temperature (Figure 12-11). Later, as the star stabilizes helium fusion in its core, its luminosity recovers, but its surface still grows hotter.

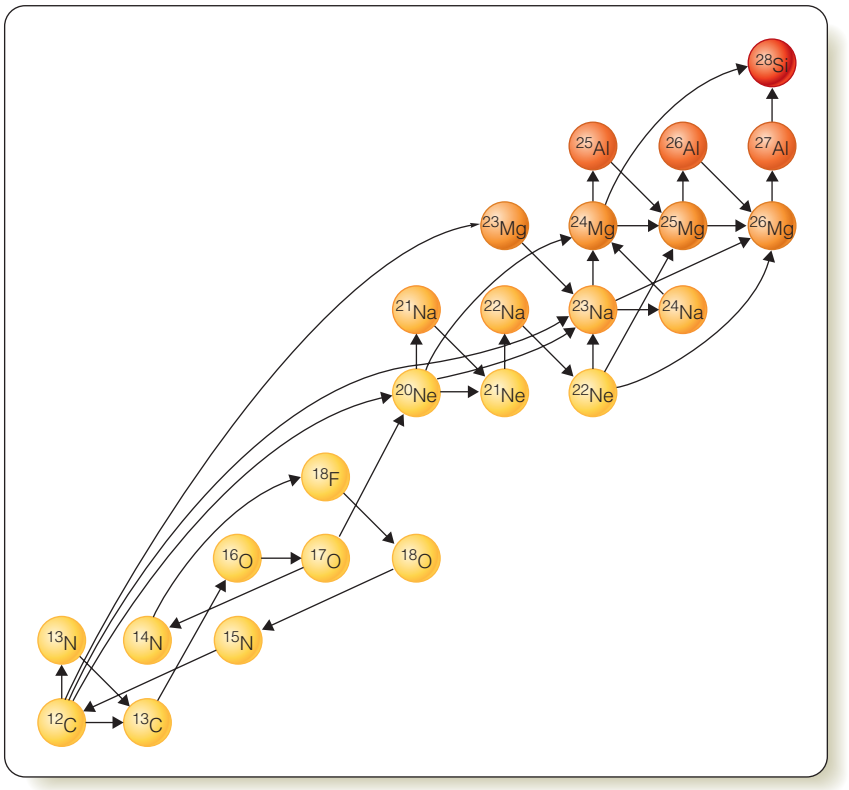
Stars fusing helium are doing something that is important to you personally. In a later chapter you will learn that the Universe began in an event called the Big Bang that produced hydrogen, helium, and small traces of elements such as lithium. Where did all the other elements come from? Elements heavier than helium, including the carbon, nitrogen, oxygen, calcium, and iron in your body, were made in stars and blasted back into space when those stars died. The processes that built the elements are called collectively **nucleosynthesis**, which means "making of nuclei."



▶ **Figure 12-11** When a main-sequence star exhausts the hydrogen in its core, it evolves rapidly to the right in the H–R diagram as it expands to become a cool giant. It then follows a looping path as it fuses helium in its core and then fuses helium in a shell. Compare with Figure 12-7. (Evolutionary track adapted from the work of Icko Iben.)

Nucleosynthesis begins as helium fusion produces carbon, and some of the carbon nuclei absorb helium nuclei to form oxygen. A few of the oxygen nuclei absorb more helium nuclei to form neon and then magnesium. Some of these reactions release neutrons, which, having no charge, are more easily absorbed by nuclei to gradually build even heavier nuclei. These reactions involving free neutrons are not important as energy producers during helium fusion. But these slow-cooker processes form some amounts of heavier elements right up to bismuth, nearly four times heavier than iron. The following is one of the most important points to understand in an entire astronomy course: *Most of the atomic elements, including those in your body, were cooked up inside stars by these nuclear processes.*

Helium fusion produces nuclear "ashes" mostly of carbon and oxygen. Carbon and oxygen thus accumulate in the core of a medium-mass star that never gets hot enough to go on and fuse those elements at a significant rate. As the ashes accumulate, the core contracts and grows hot enough to ignite a helium-fusion shell. At this stage, such a star has two shells producing energy: a hydrogen-fusion shell with an outer boundary that continues to move outward into unburned hydrogen, leaving behind helium ash; and a helium-fusion shell that progresses outward into the helium, leaving behind carbon-oxygen ash. As the star's resulting carbon-oxygen core contracts, the outer layers of the star expand, and its point in the H–R diagram moves back up and toward the right, completing a second loop into the red giant zone, as you can see in Figure 12-11.



▲**Figure 12-12** Energy generation in giant stars more massive than about 4 $M_\odot$ begins with carbon fusion and leads to many reactions involving heavier nuclear fuels.

12-2d Fusing Elements Heavier than Helium

Understanding the evolution of stars after helium exhaustion requires highly complex modeling programs running on supercomputers, with results that are still uncertain and subject to revision. However, it is safe to say that what happens in these stages depends on the mass of the star. Only in stars that begin their lives with more than about 8 $M_\odot$ can a contracting core of carbon-oxygen ash reach temperatures as high as 600 million K that is required to fuse carbon. Subsequently, some of those stars can fuse neon, oxygen, and silicon.

Carbon fusion leads to a complex network of reactions illustrated in **Figure 12-12**, where each circle represents a possible nucleus and each arrow represents a different nuclear reaction. Nuclei can react by capturing a proton, capturing a neutron, capturing a helium nucleus, or by combining directly with other nuclei. Unstable nuclei can decay by ejecting an electron, ejecting a positron, ejecting a helium nucleus, or splitting into fragments. Slow-cooking reactions that build some atomic elements but don't generate significant amounts of energy are not shown in Figure 12-12.

As you can tell from all the arrows in Figure 12-12, the fusion of elements heavier than helium is not simple. Astronomers who model the structure and evolution of massive stars must use sophisticated nuclear physics to compute the amount of energy produced and follow the changes in chemical composition that occur as the stars fuse carbon and heavier elements. Nevertheless, one simple rule is clear: Nuclear reactions involving larger nuclei with higher Coulomb barriers require higher ignition temperatures that can only occur in higher-mass stars (**Table 12-3**).

The time a star spends as a giant or supergiant is small compared with its life on the main sequence. The Sun, for example, will spend about 11 billion years as a main-sequence star but only about a billion years as a giant. The more massive stars pass through the giant stage even more rapidly. Because of the short time stars spend as giants, you see relatively few giants in the sky even though they are very luminous. This illustrates an important principle in astronomy: The shorter the time a given evolutionary stage takes, the less likely you are to see stars in that particular stage. This explains why you see a great many main-sequence stars but few giants (see Figure 3 in the Chapter 9 Concept Art: The Family of Stars).

Eventually, all stars face collapse. No matter what the star's mass, it will eventually run out of usable fuels, and gravity will win the struggle with pressure. You will explore the ultimate deaths of stars in the next chapter.

Table 12-3 Nuclear Reactions in Massive Stars

Nuclear Fuel	Nuclear Products	Minimum Ignition Temperature	Initial Main-Sequence Mass Needed to Ignite Fusion*
H	He	4×10^6 K	0.1 $M_\odot$
He	C, O	120×10^6 K	0.5 $M_\odot$
C	N, O, Ne, Na, Mg	0.6×10^6 K	~8 $M_\odot$
Ne	O, Na, Mg	1.2×10^6 K	~10 $M_\odot$
O	Si, S, P	1.5×10^6 K	~11 $M_\odot$
Si	Ni to Fe	2.7×10^6 K	~11 $M_\odot$

* The ~ symbol means "approximately."

12-3 Star Clusters: Evidence of Stellar Evolution

The theory of stellar evolution is so complex and involves so many assumptions that astronomers would have little confidence in it, if not for evidence that is provided especially by clusters of stars. Observing and comparing the properties of star clusters of different ages lets you see clear evidence of the evolution of stars.

To grasp the difficulty of understanding stellar evolution, consider this analogy: Suppose a visitor to Earth who had never seen a tree wandered through a forest for an hour looking at mature trees, fallen seeds, young saplings, rotting logs, and rising sprouts. Could that visitor understand the life cycle of trees after only an hour of observation?

Astronomers face a similar problem when they try to understand the life story of the stars. Humans do not live long enough to see stars evolve. We see only a momentary glimpse of the Universe as it appears during the time span of human civilization, a mere snapshot in which all stages in the life cycles of stars are represented. Unscrambling these stages and putting them in order is a difficult task. Only by looking at selected groups of stars—star clusters—can patterns be seen.

12-3a Observing Star Clusters

Plotting the properties of each star in a star cluster on an H–R diagram freezes a moment in the cluster's history and makes the evolution of the stars visible. All the stars in a cluster are understood to have formed at about the same time from the same cloud of gas, so they are all about the same age and have the same chemical composition. The differences you see among stars in a cluster presumably arise only from differences in mass, and that is what makes stellar evolution visible. Study **Star Clusters and Stellar Evolution** and notice three important points and four new terms:

1. There are two kinds of star clusters, *open clusters* and *globular clusters*. They look different, but their stars evolve in similar ways. You will learn more in a later chapter about how these clusters help astronomers understand the history of our Galaxy.
2. You can estimate the age of a star cluster by noting the *turnoff point* in the distribution of data points that represent its stars in the H–R diagram. The turnoff point represents the hottest and most luminous stars remaining on the main sequence.
3. Finally, notice that the shape of a star cluster's H–R diagram is governed by the evolutionary paths the stars take. The H–R diagrams of older clusters are especially clear in outlining how stars evolve away from the main sequence to the giant region and then move left along the *horizontal branch* before turning back toward the giant region. By comparing clusters of different ages, you can visualize how stars evolve almost as if you were watching a film of a star cluster evolving over billions of years.

Were it not for star clusters, astronomers would have little confidence in the theories of stellar evolution. Star clusters make that evolution visible and assure astronomers that they really do understand most of how stars are born, live, and die.

12-3b The Evolution of Star Clusters

A star cluster is formed when a cloud of gas contracts, fragments, and forms a group of stars. As you recall from the previous chapter regarding star formation, some groups of stars, called *associations*, are so widely scattered that the group is not held together by its own gravity. The stars in an association wander away from each other rather quickly. Some groups of stars are more compact; even after the stars become luminous and the gas and dust are blown away, gravity holds the group together as a star cluster.

Open clusters are not as old or as crowded as the globular clusters, and that helps explain their appearance. Close encounters between stars are rare in an open cluster where the stars are farther apart, and such clusters have an irregular appearance. The globular clusters appear to be nearly perfect globes because the stars are much closer together and encounters between stars are more common. The globular clusters have had time to evenly distribute the energy of motion among all of the stars, so they have settled into a more uniform, spherical shape.

As a star cluster ages, some stars traveling a bit faster than the rest can escape. Globular clusters are compact and massive; most have survived for 11 billion years or more. In contrast, a star cluster with only a few widely distributed stars may evaporate completely as its stars escape one by one. Our Sun may have formed in a star cluster 5 billion years ago. Some astronomers have pointed out that the open cluster M67 has almost exactly the same composition and age as the Sun and wondered if the Sun is a lost member of that cluster. The motions of M67 and the Sun through the Galaxy are quite different, however, so it is hard to imagine how they could really be associated. There is probably no chance of tracing the Sun back to its original family home, even if that cluster still exists.

Practicing SCIENCE

What evidence can you cite that giant stars are main-sequence stars that expanded to large diameters? A scientist answering this question must combine multiple sets of observations, considering both the H–R diagram and the main-sequence mass–luminosity relationship.

You know that main-sequence stars have a mass–luminosity relation, a mathematical formula expressing the fact that more massive stars are also more luminous. You also know the reason for that relation lies with the pressure–temperature thermostat: More massive stars must be hotter in their centers to be stable. Further, you know that higher temperatures mean faster fusion reaction rates and therefore faster consumption of hydrogen fuel and that high-mass stars actually have shorter lifetimes than low-mass stars.

Armed with that knowledge, when you examine H–R diagrams of star clusters, you notice that clusters that contain all types of main-sequence stars, from low-mass M dwarfs to high-mass O stars, do not have red giants or supergiants. You also notice, for example, that clusters with missing O, B, and A main-sequence stars do have red giants and supergiants. Aha—you realize that, in clusters older than the main-sequence lifetime of O, B, and A stars, stars of those masses have left the main sequence and become red giants and supergiants. Only F, G, K, and M stars remain on the main sequence. The former O, B, and A main-sequence stars have now run out of nuclear fuel in their cores, expanded, and become red giants and supergiants. The pace of stellar evolution is displayed in a cluster H–R diagram as clearly as a map.

12-4 Variable Stars: Evidence of Stellar Evolution

It is a **Common Misconception** that the stars are constant and unchanging. Stellar models show that stars slowly evolve as they consume their fuels, and evidence from H–R diagrams of main-sequence stars and star clusters confirms these slow changes in structure. But is that enough evidence to be certain? One of the most important principles of science is that hypotheses must be checked against all known evidence. Some stars that vary in brightness provide evidence of stellar evolution and illustrate the importance of the way in which energy flows outward through the interiors of the stars.

A **variable star** is any star that changes its brightness significantly and repeatedly. Some variable stars are eclipsing binaries (look back to Section 9-5c), but many others are single stars that grow brighter and fainter because of internal processes. These stars are often called **intrinsic variables** to distinguish them from eclipsing binaries. Of these intrinsic variables, two related kinds are centrally important to modern astronomy, even though the first star of that class was discovered centuries ago.

12-4a Cepheid and RR Lyrae Variable Stars

In 1784, the deaf and mute English astronomer John Goodricke, then 19 years old, discovered that the star Delta Cephei is variable, changing its brightness by almost a magnitude over a period of 5.37 days (**Figure 12-13a**). Goodricke died at the age of 21, but his discovery ensures his place in the history of astronomy. Since 1784, hundreds of stars like Delta Cephei have been found, and they are now known as **Cepheid variable stars**.

Cepheid variables are supergiant or bright giant stars of spectral type F or G. The fastest of them complete a variation cycle—from bright to faint to bright again—in about 2 days, whereas the slowest take as long as 60 days. A plot of a variable's magnitude versus time, known as a light curve (for example, Figures 9-20 and 9-21), shows a typical rapid rise to maximum brightness and a slower decline (**Figure 12-13b**). Some Cepheids change their brightness by only 0.01 magnitude (about 1 percent), whereas others change by more than a magnitude. One of the Famous Stars, Polaris—the North Star—is a Cepheid with a period of 3.97 days and an amplitude of less than 5 percent. The **RR Lyrae variables**, a related type of star, are giant stars that are fainter than the Cepheids and pulsate with periods of less than a day.

Studies of Cepheids and RR Lyrae stars reveal two related facts of great importance. First, there is a **period–luminosity relation** that connects the period of pulsation of a Cepheid to its luminosity. The longer-period Cepheids are about 40,000 times more luminous than the Sun, whereas the shortest-period Cepheids are only a few hundred times more luminous than the Sun. Second, there are two types of Cepheids. Type I Cepheids, including Delta

Star Clusters and Stellar Evolution

1 An **open cluster** is a collection of 10 to 1000 stars in a region about 25 pc in diameter. Some open clusters are quite small, and some are large, but they all have an open, transparent appearance because the stars are not crowded together.

In a star cluster, each star follows its orbit around the center of mass of the cluster.

NOAO/AURA/NSF

Open Cluster
The Jewel Box

Visual

1a A **globular cluster** can contain 10^5 to 10^6 stars in a region only 10 to 30 pc in diameter. The term *globular cluster* comes from the word *globe*, although globular cluster is pronounced like *glob* as in glob of butter. These clusters are nearly spherical, and the stars are much closer together than the stars in an open cluster.

Anglo-Australian Observatory / David Malin Images

Astronomers can construct an H–R diagram for a star cluster by plotting a point to represent the luminosity and temperature of each star.

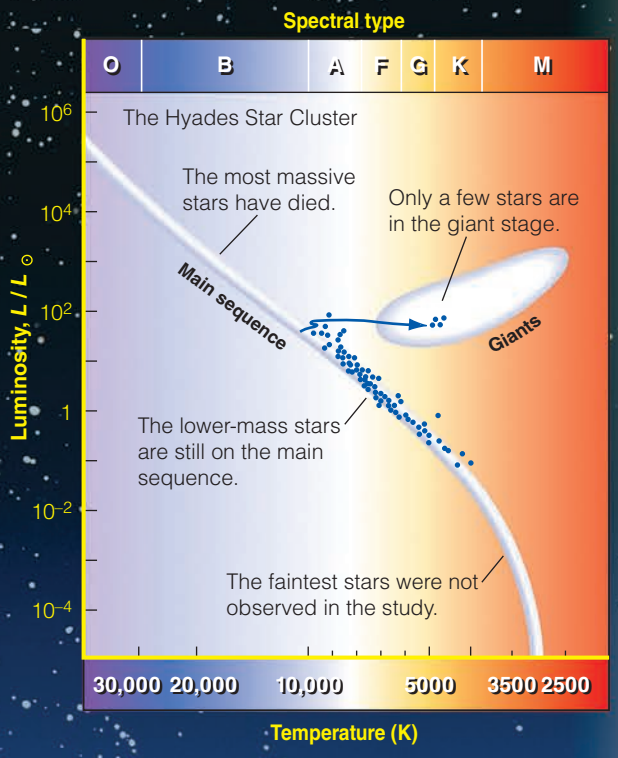


Illustration design by Michael A. Seeds

2 The H–R diagram of a star cluster can make the evolution of stars visible. The key is to remember that all of the stars in the star cluster have the same age and composition, differing only in mass. The H–R diagram of a star cluster provides a "snapshot" of the evolutionary state of the stars at the time you happen to observe them. This diagram shows the 650-million-year-old star cluster called the Hyades, easily visible to the unaided eye in the constellation Taurus. The upper main sequence is missing because the more massive stars have died, and this snapshot catches a few medium-mass stars leaving the main sequence to become giants.

As a star cluster ages, its main sequence grows shorter, like a candle burning down. You can judge the age of a star cluster by looking at the turnoff point, the point on the main sequence where stars are currently evolving to the right to become giants. Stars at the **turnoff point** have lived out their lives and are about to die. Consequently, the life expectancy of the stars at the turnoff point equals the age of the cluster. (Evolutionary track adapted from the work of Icko Iben.)

3 From theoretical models of stars, you could construct a film to show how the H–R diagram of a star cluster changes as it ages. You can then compare theory (*left*) with observation (*right*) to understand how stars evolve. Note that the time step for each frame in this film increases by a factor of 10.

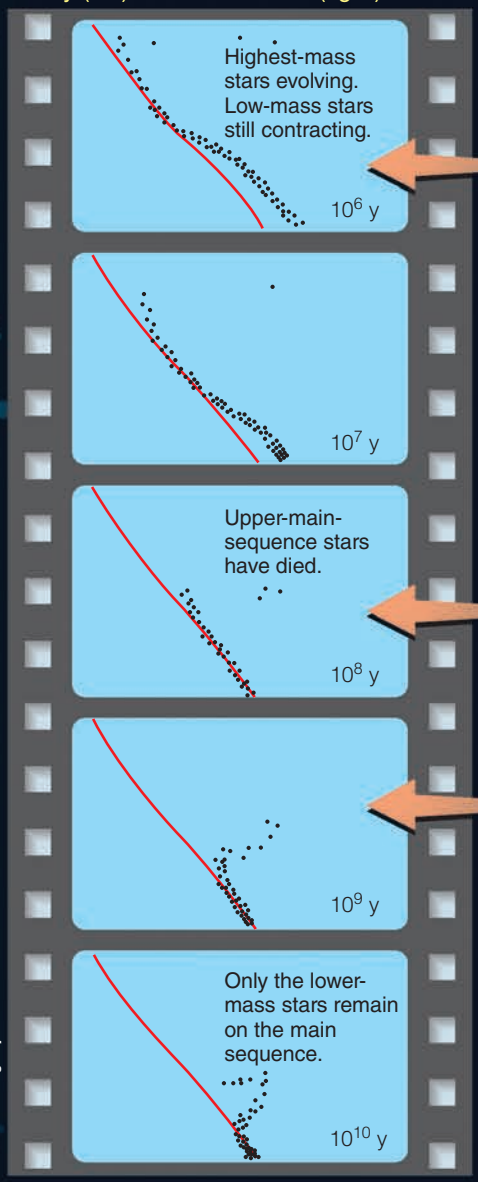
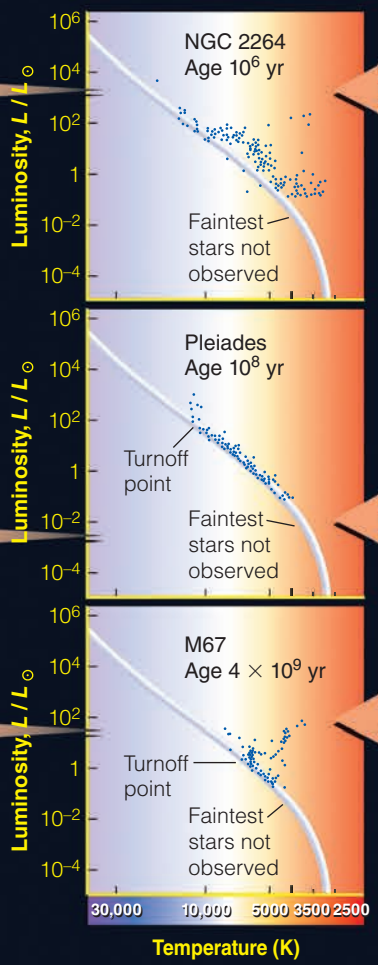
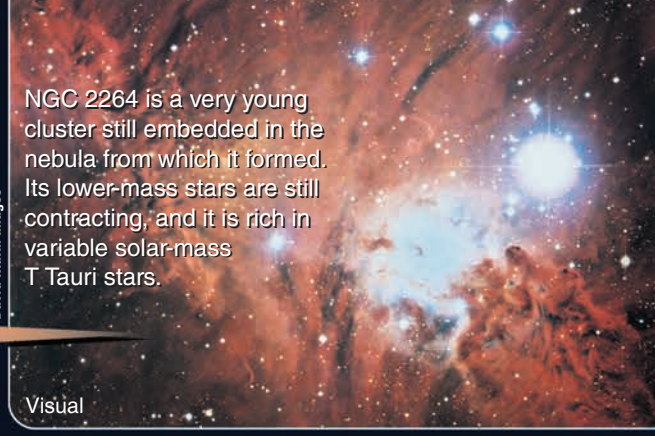


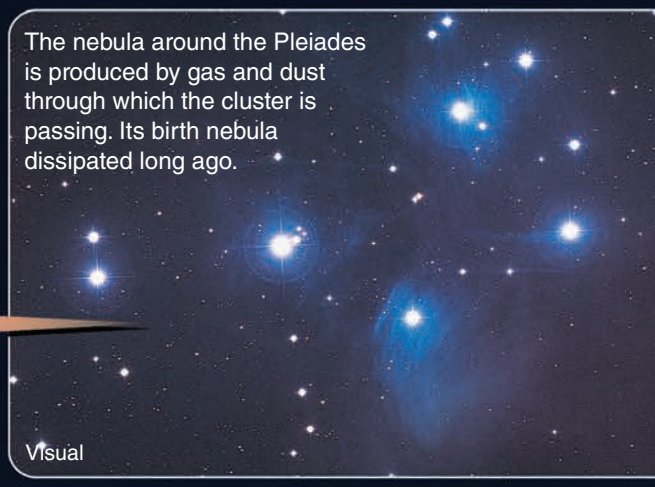
Illustration design by Michael A. Seeds





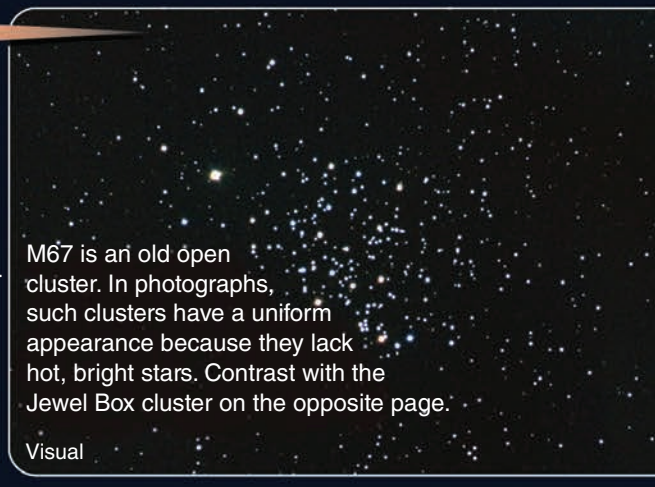
Anglo-Australian Observatory/ David Malin Images

NGC 2264 is a very young cluster still embedded in the nebula from which it formed. Its lower-mass stars are still contracting, and it is rich in variable solar-mass T Tauri stars.



Caltech

The nebula around the Pleiades is produced by gas and dust through which the cluster is passing. Its birth nebula dissipated long ago.



NOAO/AURA/NSF/N. Sharp, M. Hanna

M67 is an old open cluster. In photographs, such clusters have a uniform appearance because they lack hot, bright stars. Contrast with the Jewel Box cluster on the opposite page.

Globular cluster H–R diagrams resemble the last frame in the film, which tells you that globular clusters are very old.

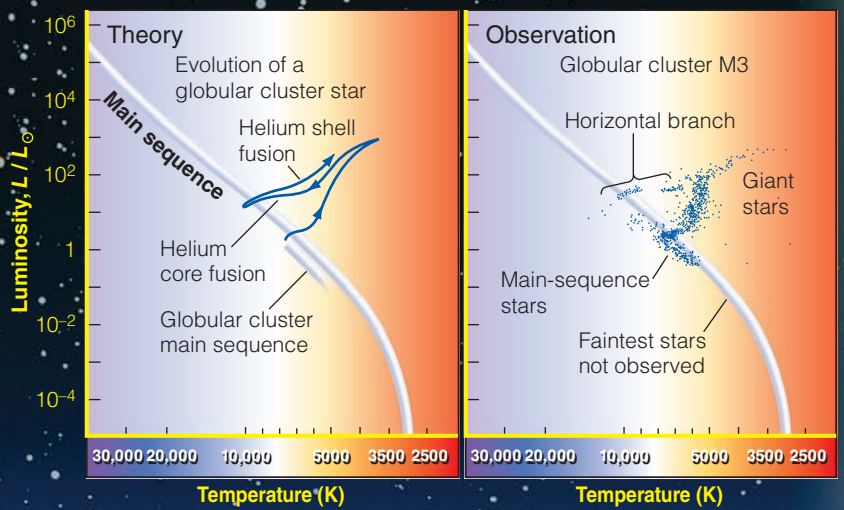
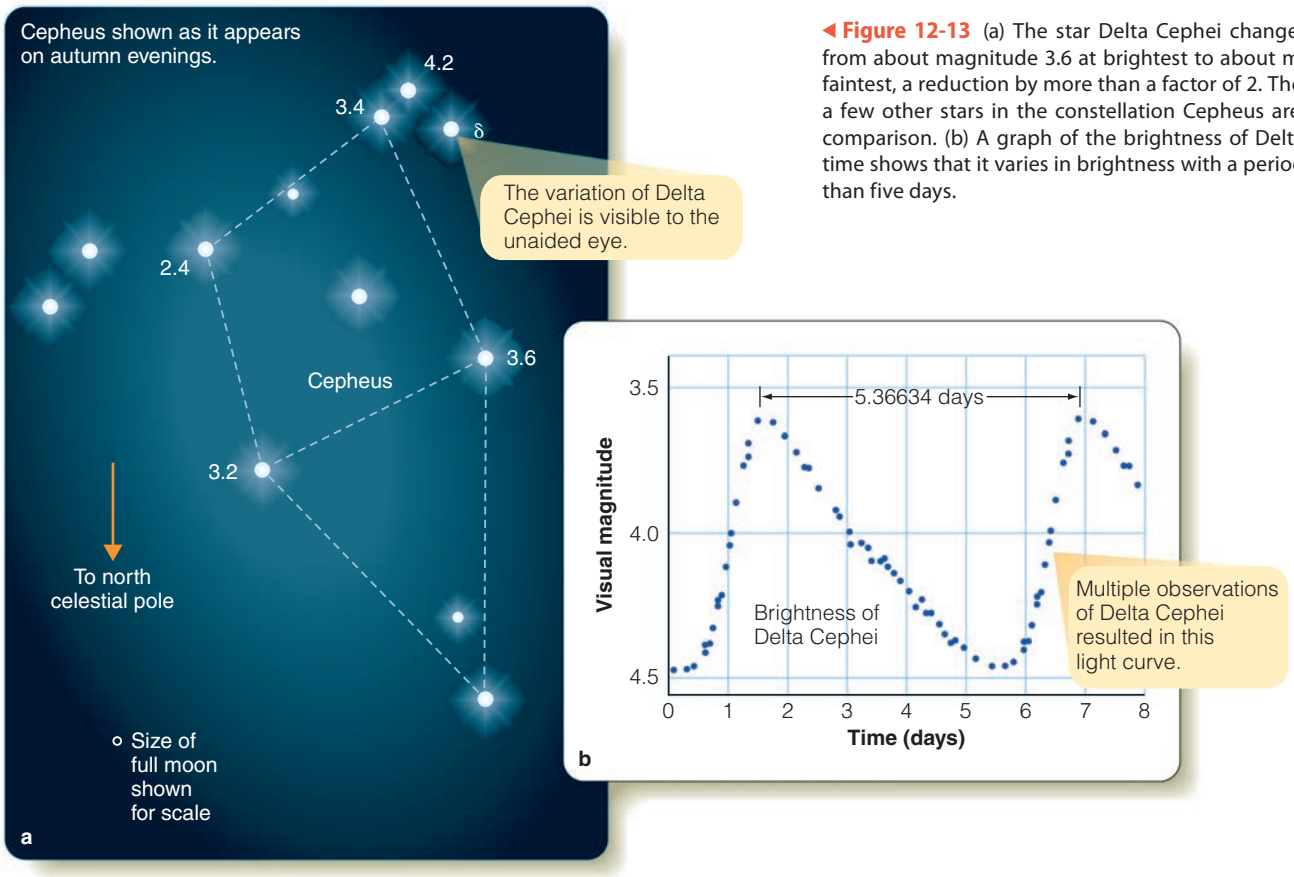


Illustration design by Michael A. Seeds

3a The H–R diagrams of globular clusters have very faint (low-luminosity, low-mass) turnoff points showing that they are very old clusters. The most complete analysis suggests these clusters are about 11 billion years old.

The **horizontal branch** stars are giants fusing helium in their cores or in shells. The shape of the horizontal branch outlines the evolution of these stars.

The main-sequence stars in globular clusters are fainter and bluer than the zero-age main sequence. Spectra reveal that globular cluster stars are poor in elements heavier than helium, which means their gases are less opaque. That means energy can flow outward more easily, making these low-metallicity stars slightly smaller and hotter than their high-metallicity counterparts with the same masses.



◀ **Figure 12-13** (a) The star Delta Cephei changes its brightness from about magnitude 3.6 at brightest to about magnitude 4.5 at faintest, a reduction by more than a factor of 2. The magnitudes of a few other stars in the constellation Cepheus are given here for comparison. (b) A graph of the brightness of Delta Cephei versus time shows that it varies in brightness with a period slightly longer than five days.

Cephei itself, have chemical compositions like that of the Sun, but type II Cepheids and the RR Lyrae stars are poor in elements heavier than helium. These facts are clues that will help you understand variable stars as evidence of stellar evolution.

12-4b Pulsating Stars

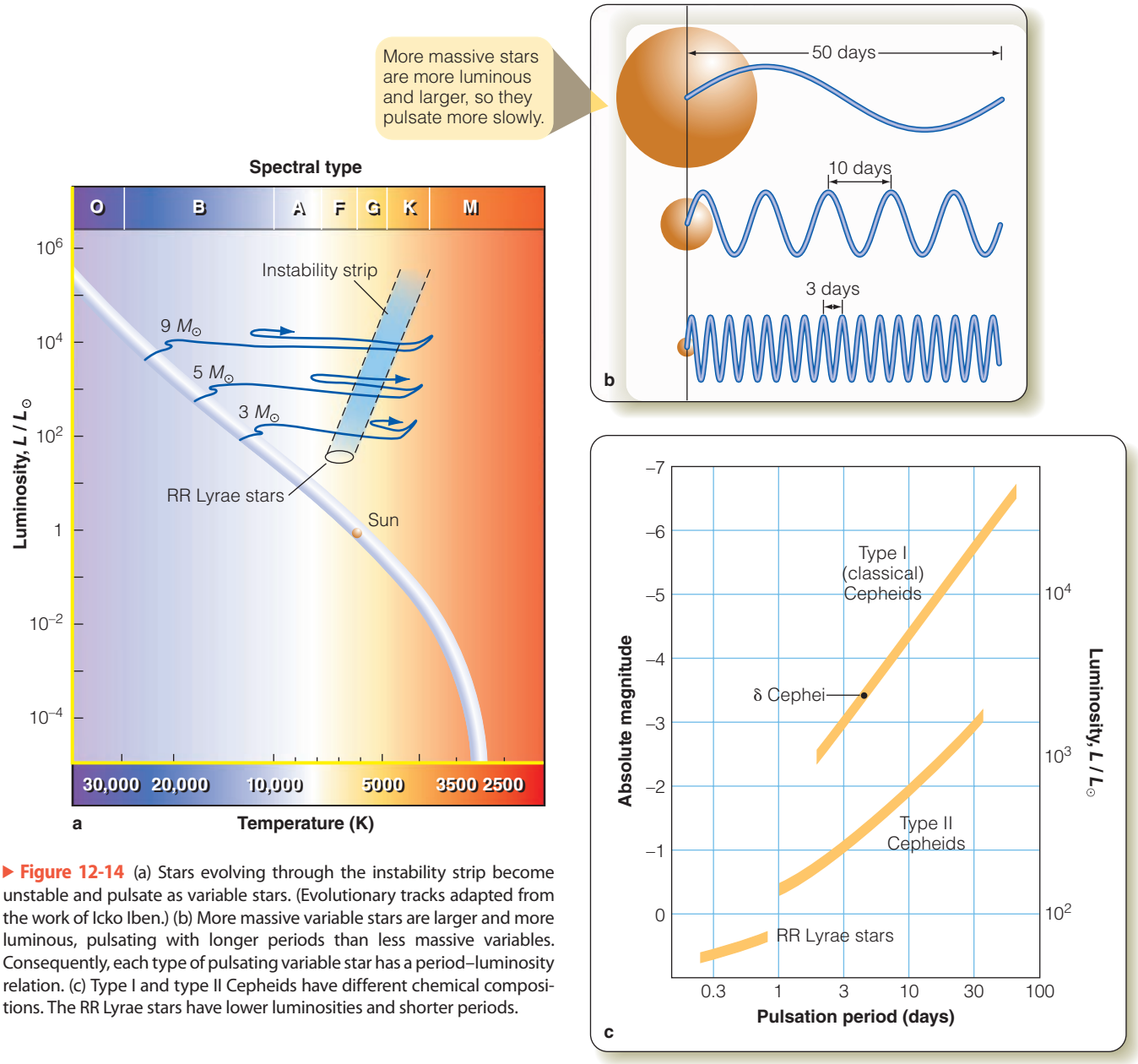
Why would a star pulsate? Why should there be a period–luminosity relation? Why are there two types of Cepheids? The answers to these questions will reveal some of the deepest secrets of the stars. To find the answers, you should start by reconsidering the H–R diagram.

Remember that after a star leaves the main sequence, it fuses hydrogen in a shell, and the point that represents it in the H–R diagram moves to the right as the star becomes a cool giant. When it ignites helium in its core, the star contracts and grows hotter, and the point in the diagram moves to the left. Soon, however, the helium core is exhausted; helium fusion continues only in a shell, forcing the star to expand again, and the point in the H–R diagram moves back to the right, completing a loop. As giant stars evolve through these stages, their sizes and temperatures change in complicated ways. Certain combinations of size and temperature make a star unstable, causing it to pulsate as an intrinsic variable star. In the H–R diagram, the region where size and temperature lead to pulsation is called the **instability strip**, as shown in the H–R diagram in **Figure 12-14a**.

What makes variable stars pulsate? Start with a stable star. You could make a stable star pulsate temporarily if you squeezed it and then released it. The star would rebound outward, and it might oscillate in and out for a few cycles, but the oscillation would eventually run down because of friction. The energy of the moving gas would be converted to heat and radiated away, and the star would return to stability with each part of its interior in hydrostatic equilibrium. To make a star pulsate continuously, some process must drive the oscillation, just as a spring in a windup clock is needed to keep it ticking.

Studies using stellar models show that variable stars in the instability strip pulsate like beating hearts because of an energy-absorbing layer in their outer envelopes. This layer is the region where helium is partially ionized. Above this layer, the temperature is too low to ionize helium, and below the layer it is hot enough to ionize helium completely. Like a spring, the helium ionization zone can absorb energy when it is compressed and release it when the zone expands, and that is enough to keep the star pulsating.

You can follow this pulsation in your imagination. As the outer layers of a pulsating star expand, the ionization zone expands, and the ionized helium becomes less ionized and releases stored energy, which gives the expanding gas a little boost and the expansion goes a little faster. The surface of the star expands too far—overshoots its equilibrium position—and



▶ **Figure 12-14** (a) Stars evolving through the instability strip become unstable and pulsate as variable stars. (Evolutionary tracks adapted from the work of Icko Iben.) (b) More massive variable stars are larger and more luminous, pulsating with longer periods than less massive variables. Consequently, each type of pulsating variable star has a period–luminosity relation. (c) Type I and type II Cepheids have different chemical compositions. The RR Lyrae stars have lower luminosities and shorter periods.

eventually falls back. As the surface layers contract, the ionization zone is compressed and becomes more ionized, which absorbs energy. Robbed of some of their energy, the layers of the star can't support the weight, and they compress a little faster. This allows the contraction to go even faster, so the infalling layers overshoot that equilibrium point until the pressure inside the star slows them to a stop and makes them expand again.

Cepheids change their radii by 5 to 10 percent as they pulsate, and this motion can be detected by observing Doppler shifts in their spectra. When the stars expand, the surface layers approach us, and astronomers detect a blueshift. When they contract, the surface layers recede, and astronomers detect a redshift. Although a 10 percent change in radius seems like a large change, it affects only the outer layers of the star. The center of the star is much too dense to participate in the pulsations.

Only stars in the instability strip have their helium ionization zone at the right depth to act as a spring and drive the pulsation. In cool stars, the zone is too deep and cannot overcome the weight of the layers above it. That is, the "spring" is overloaded and can't expand. In hot stars, the helium ionization zone lies near the surface, and there is little weight above it to compress it. In that case, the "spring" never gets squeezed. Stars in the instability strip have the right combination of temperature and radius for the helium ionization zone to fall exactly where it is most effective at driving pulsation.

Now you can also understand why there is a period–luminosity relation. The more massive stars are larger and more

luminous. As they evolve and cross into the H–R diagram's instability strip, they are higher in the diagram. But the larger a star is, the more slowly it pulsates (**Figure 12.14b**)—just as large bells vibrate slowly and go "bong," whereas little bells vibrate rapidly and go "ding." Pulsation period depends on size, which depends on mass, and mass also determines luminosity. Therefore, the most luminous Cepheids at the top of the instability strip are also the largest and pulsate most slowly. Less luminous Cepheids are smaller and pulsate faster. If you plot the average brightness of the variable stars against their periods of pulsation, you can clearly see the relationship between luminosity and period. Look at **Figure 12-14c** to see this relationship.

You know enough about stars to understand why there are two types of Cepheid variable stars. Type I Cepheids have chemical abundances roughly like those of the Sun, but type II Cepheids are poor in elements heavier than helium. That means the gases in type II Cepheids are not as opaque as the gases in type I Cepheids. The energy flowing outward can escape more easily in type II Cepheids, and they reach a slightly different equilibrium. For stars of the same period of pulsation, type II Cepheids are less luminous, as you can see in Figure 12-14c.

In the H–R diagram, the RR Lyrae stars lie at the lower end of the instability strip. They are a bit smaller and hotter than Cepheids, but they pulsate for the same reason.

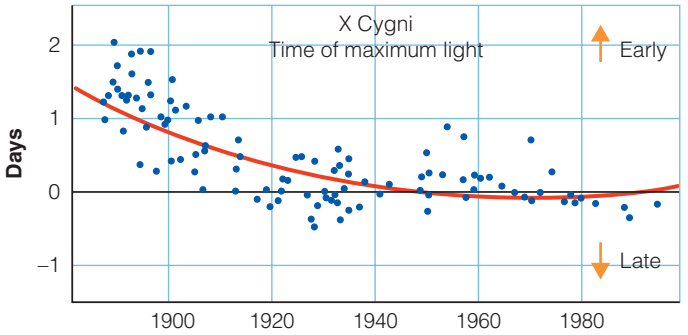
12-4c Period Changes in Variable Stars

Some Cepheids have been observed to change their periods of pulsation, and those changes are clear evidence that stars really do evolve.

The evolution of a star may carry it through the instability strip a number of times, and each time it can become a variable star. If you could watch a star as it entered the instability strip, you could see it begin to pulsate. Similarly, if you could watch a star leave the instability strip, you would see it stop pulsating. Such an observation would be dramatic evidence that the stars are evolving, but stars evolve so slowly that these events are rare.

Clear evidence of stellar evolution can be found in the slowly changing periods of Cepheid variables. Even a tiny change in period can become easily observable because it accumulates, just as a clock that gains a second each day becomes obviously fast over as the course of a year. Some Cepheids have periods that are growing shorter, and others have periods that are growing longer. These changes occur because the stars are evolving and changing their average radii; contraction shortens the period, and expansion lengthens the period. The star X Cygni, for example, has a period that is gradually growing longer (**Figure 12-15**). These changes in the periods of pulsating stars provide definite evidence that the stars are changing their internal structures and evolving.

One famous Cepheid, RU Camelopardalis (RU Cam for short), was observed to stop pulsating in 1966. Most astronomers assumed that it had stopped pulsating because it was leaving the instability strip, but RU Cam's pulsations resumed in 1967, so that star's exact situation is not well understood. The pulsation amplitude of Famous Star Polaris has been decreasing during the past few decades, and that star may be going through a similar change now, although it does not lie near the edge of the instability strip in the H–R diagram. There may be other undiscovered factors affecting the evolution of Cepheids like RU Cam and Polaris.



▲ **Figure 12-15** Like a clock running just a bit slow, the Cepheid variable star X Cygni has been reaching maximum brightness later and later for most of this century. This is shown by the downward curve in this graph of its predicted minus observed times of maximum brightness. As it evolves to the right across the instability strip, it is slowly expanding, and its period is growing longer by 1.46 seconds per year.

There are other kinds of variable stars that are not in the H–R diagram's instability strip because their pulsations are driven by different mechanisms than Cepheids. Only Cepheids are discussed here because they are relatively common, they illustrate important ideas about stellar structure, and their changing periods make stellar evolution clearly evident. Also, you will use Cepheids in later chapters to study the distances to galaxies and the expansion of the Universe.

Practicing SCIENCE

What evidence can you cite to show that stars really are evolving? Some processes occur so slowly that they cannot be observed over a human lifetime, so scientists must infer those processes from short-term observations.

Stellar evolution occurs over billions of years, but because astronomers can plot H–R diagrams of star clusters that have different ages, they can see how stellar evolution affects the stars. Those H–R diagrams can be assembled into a sequence that makes stellar evolution obvious. Variable stars provide further evidence of stellar evolution because small changes in radius as their internal structures change produce detectable changes in the pulsation periods. Because those changes accumulate and make the stars resemble clocks that run "fast" or "slow," astronomers can observe the effect of stellar evolution.

Now, as scientists do, consider the relationship between models and observations: ***How do stellar models allow astronomers to understand why Cepheid and RR Lyrae stars pulsate?***

WHAT ARE WE?

Snap-Shooters

A human lifetime is less than a century, and that is a mere flicker of a moment in the history of the Universe. What we see in the sky during our lives is just a snapshot that freezes the action. The stars form gradually and evolve over millions and billions of years. We humans see none of that action. Our snapshot shows us the many stages of star birth and evolution, but we see almost none of the changes through which stars pass.

Look again at Section 1-2, "When Is Now?" The entire 10,000-year history of human civilization is only a tiny fraction of the history of the Universe. In that time, a few of the most massive stars have been observed to evolve slightly, but the vast majority of stars have not changed since humans on Earth began building the first cities.

Only by the application of human ingenuity have astronomers figured out how stars work, how they are born, how they evolve, and, as you will see in the next chapter, how they die. Our human life spans allow us only to make snapshots, but they reveal that we are part of a complex and evolving Universe.

Study and Review

SUMMARY

12-1 Stellar Models

- **Stellar models** of the interiors of stars show the evolution of stars based on four simple laws of stellar structure. The first, the law of hydrostatic equilibrium, says that the star must balance the weight of its outer layers by its internal pressure. The second law is energy transport, which says that energy can flow from hot to cooler regions only by conduction, convection, or radiation. The final two laws are the **law of mass conservation** and the **law of energy conservation**.
- The mass–luminosity relation for main-sequence stars is explained by the requirement that a main-sequence star support the weight of its layers by its internal pressure. The more massive a star is, the more weight the star must support, and the higher the star's internal pressure must be. To keep the star's pressure high, the star must be hot and generate large amounts of energy. Thus, the mass of a main-sequence star determines the star's luminosity.
- Extremely massive stars are quite rare. Contracting gas clouds tend to fragment and produce pairs, groups, or clusters of stars, not many of which are extremely massive stars. Also, massive stars tend to blow matter away in strong stellar winds, which rapidly reduces their mass.
- Objects less massive than 0.08 solar mass can't get hot enough to sustain hydrogen fusion. Thus they become **brown dwarfs**, rather than normal stars, and slowly cool as they radiate away their thermal energy. Many brown dwarfs have been observed.
- The **zero-age main sequence (ZAMS)** is the line in the H–R diagram representing the properties of contracting stars that begin fusing hydrogen and reach hydrostatic equilibrium. While a star is on the main sequence, it moves slightly upward and to the right in the H–R diagram. Thus, the main sequence is a band rather than a narrow line.
- The life expectancy of a main-sequence star depends on the star's mass.

$$t_* = \frac{M_*}{L_*} = \frac{M_*}{M_*^{3.5}} = \frac{1}{M_*^{2.5}} \qquad \text{(Eq. 12-1)}$$

More massive stars have more fuel but use the fuel very rapidly and thus remain on the main sequence of the H–R diagram only a few million years. The Sun, a medium-mass star, will last a total of about 11 billion years on the main sequence. The least massive stars can survive for hundreds of billions of years on the main sequence.

12-2 Post-Main-Sequence Evolution

- When a main-sequence star exhausts its hydrogen, its core contracts, and it begins to fuse hydrogen in a shell around its core. The outer part of the star—its envelope—expands and the surface of the star cools. A Sun-like main-sequence star moves toward the right in the H–R diagram to become a giant. Main-sequence stars much more massive than the Sun move across the top of the H–R diagram as supergiants.
- The core is where the nuclear fusion occurs and is small and dense. Extremely dense gas becomes **degenerate matter**, which behaves in two surprising ways. First, it resists compression, and second, its pressure does not depend on temperature.
- When the core of a giant becomes hot enough, helium fusion begins first in the core and then in a shell producing carbon via the **triple alpha process**. Some stars begin helium fusion gradually, but stars in a certain mass range begin with an explosion called a **helium flash**. Stars fusing helium build most of the chemical elements in the Universe through **nucleosynthesis**.
- An evolving main-sequence star that is more massive than the Sun has a core that can ignite carbon to produce oxygen, which is a fuel of higher mass. The core now looks like an onion, with higher-mass carbon fusion occurring in the core and lower-mass helium and hydrogen fusion occurring in nested shells. Each newly generated, higher-mass fuel fuses more rapidly, producing heavier and heavier elements in the core, depending on the star's mass, up to the element iron.

12-3 Star Clusters: Evidence of Stellar Evolution

- Because all the stars in a cluster have about the same chemical composition and are formed around the same time, effects of stellar evolution among the different mass stars of the cluster are apparent when plotted in the same H–R diagram. This plot shows the most massive stars leaving the main sequence first because massive stars evolve faster than lower-mass stars.
- There are two types of star clusters. **Open clusters** contain 10 to 1000 stars and have an open, transparent appearance. **Globular clusters** contain 10^5 to 10^6 stars densely packed into a spherical shape. The open clusters tend to be young to middle aged; thus they have a blue color. Globular clusters have ages up to 13 billion years; thus they have a red color. Also, globular clusters tend to be poorer in elements heavier than helium.

- You can determine the age of a cluster by looking at the **turnoff point**, the location of the hottest, most luminous stars remaining on the main sequence on the H–R diagram, where the stars turn off to the right and evolve into supergiants or giants. The life expectancy of a star at the turnoff point equals the age of the cluster.
- H–R diagrams of old star clusters such as globular clusters show how giant stars ignite helium fusion in their cores and evolve to the left in the diagram along the **horizontal branch**, after the helium core flash.

12-4 Variable Stars: Evidence of Stellar Evolution

- **Variable stars** are those that change in brightness significantly and repeatedly over time. Some variable stars are in binaries, but **intrinsic variable stars** change in brightness because of internal processes and not because they are members of eclipsing binaries.
- The **Cepheid variable stars** and **RR Lyrae variable stars** are intrinsic variables that pulsate. The Cepheids obey a **period–luminosity relation** because more massive stars, which are more luminous, are larger and pulsate more slowly.
- Pulsating stars have luminosities and temperatures that put them in an **instability strip** of the H–R diagram. A helium ionization layer in the star's envelope stores and releases energy as the star expands and contracts. A star outside the instability strip does not pulsate because the helium ionization layer is too deep or too shallow to make the star radially unstable.
- Some Cepheids have periods that are slowly changing, indicating that the evolving star's average radius and thus its period of pulsation are gradually changing. Thus, some intrinsic variable stars show direct evidence of stellar evolution.

REVIEW QUESTIONS

1. Why is there a main sequence?
2. What is the observed upper mass limit to the main sequence? Why is there an upper mass limit?
3. What is the observed lower mass limit to the main sequence? Why is there a lower mass limit?
4. Explain how the law of conservation of mass applies to models of the interiors of stars.
5. Explain how the law of conservation of energy applies to the interiors of stars.
6. Describe the law of hydrostatic equilibrium.
7. Describe the law of energy transport. How is energy transported inside main sequence stars having the lowest mass? The highest mass?
8. What is a brown dwarf?
9. Rank Jovian planets, stars, and brown dwarfs in order of decreasing mass.
10. Why is there a mass–luminosity relation for main-sequence stars?
11. Why does the main sequence appear as a band and not just the ZAMS (zero-age) line?
12. Why does a star's life expectancy depend on mass?
13. Which one lives longer, an O3 main-sequence star or a K7 main-sequence star? Why?
14. Star A and Star B are both on the main sequence. Star A is 80 times more luminous than Star B. Which star is more massive? How do you know? Which star has a longer life? Why?
15. Why do expanding, aging stars become cooler and more luminous?
16. Define *degeneracy* in relation to stars in your own words.
17. Which spectral types of main-sequence stars will experience a helium flash?
18. What causes the helium flash? Why does a helium flash make understanding the later stages of stellar evolution more difficult?
19. How do some stars avoid the helium flash?
20. What gives the triple-alpha process its name? Why is it called a process and not a chain or a cycle?
21. What fusion cycle or chain occurs in the core of a main-sequence star with 0.8 $M_\odot$? Which elements in the periodic table will it produce during its entire life span, both main-sequence stage and red giant stage? Which of those elements will be produced in the core, and which in shells?
22. Why do giant stars have such low overall density?
23. Why are main-sequence stars like the Sun unable to ignite more massive nuclear fuels such as carbon?
24. What fusion cycle or chain occurs in the core of a main-sequence star with 15 $M_\odot$?
25. Which elements are created in stellar nucleosynthesis?
26. How can you estimate the age of a star cluster?
27. A star cluster with 100 main-sequence stars ranging in color from red to blue has an irregular shape. Which type of star cluster is it? Is its age more like the age of M67, or the Pleiades, or NGC 2264? (Refer to the Concept Art: Star Clusters and Stellar Evolution.)
28. For a star with 1.2 $M_\odot$, put the following evolutionary stages in chronological order: ZAMS, supergiant, T Tauri, horizontal branch, main sequence, protostar, giant, and helium flash. (*Note:* All stages might not be used.)
29. How do star clusters confirm that stars evolve?
30. Will the Sun pass through the instability strip? Why or why not?
31. How do some variable stars confirm that stars have evolved?
32. **How Do We Know?** How can mathematical models allow scientists to study processes that are hidden from human eyes or happen too fast or too slowly for humans to experience?

ACTIVE INQUIRY QUESTIONS

1. How do you know that the helium flash occurs if the flash cannot be observed? Can you accept an event as real if you can never observe it?
2. Can you think of ways that chemical differences could arise among stars in a single star cluster? Consider the mechanism that triggered their formation.
3. Some stars pulsate or cyclically change radius as they evolve off the main sequence. Will the Sun experience this pulsation in the future? If so, what will happen to the planets?

PROBLEMS

1. In the model shown in Figure 12-2, what fraction of the Sun's mass is hotter than 13 million K?
2. Using the model shown in Figure 12-2, estimate the fraction of the Sun's radius that is the hydrogen-fusion core. How much denser is the center of the Sun than the layer 90 percent of the way from the center to the surface? How much hotter is the center of the Sun than the surface of the Sun?
3. What is the life expectancy of a 16-$M_\odot$ main-sequence star? Of a 50-$M_\odot$ main-sequence star?
4. Star A and Star B are both on the main sequence. Star A is 80 times more luminous than Star B. What is the ratio of their main-sequence lifetimes? (*Hint*: Refer to the stellar life expectancies equation, Eq. 12-1.)
5. What is the mass of a star with a main-sequence lifetime of 5 billion years?

6. Star A is a 150-$M_\odot$ main-sequence star, and Star B is a 0.08-$M_\odot$ main-sequence star. How much longer or shorter will Star A live compared to the Sun? What about Star B? How much more or less luminous is Star A compared to the Sun? What about Star B?
7. Compute how long the lowest-mass main-sequence star will live. Compare this age to the age of the Universe as given in Section 1-2. Which stage is this star (for example, protostar, ZAMS, main sequence, giant, supergiant, and so on) in its evolution?
8. If the Sun expanded to a radius 100 times its present radius, what would its average density be? (*Note*: The volume of a sphere is $\frac{4}{3}\pi r^3$.)
9. If a giant star 100 times the diameter of the Sun is 1 parsec (pc) from Earth, what would its angular diameter be? (*Hint*: Use the small-angle formula, Eq. 3-1.) (*Notes:* The diameter of the Sun can be found in Celestial Profile: The Sun in Chapter 8; 1 pc $= 3.1 \times 10^{13}$ km.)
10. What fraction of the volume is the helium core in a giant star of 5 $M_\odot$? (*Hint:* See Figure 12-9.) (*Note*: The volume of a sphere is $\frac{4}{3}\pi r^3$.)
11. If the stars at the turnoff point in a star cluster have masses of about 4 $M_\odot$, how old is the cluster? (*Hint*: Refer to the stellar life expectancies equation, Eq. 12-1.)
12. If an open cluster contains 500 stars and is 25 pc in diameter, what is the average distance between the stars? (*Hint*: On average, what share of the volume of the cluster surrounds each star?) (*Note*: The volume of a sphere is $\frac{4}{3}\pi r^3$.)
13. Repeat the previous problem for a typical globular cluster containing a million stars in a sphere 25 pc in diameter.
14. Calculate the distance to Delta Cephei using Figure 12-13b, Figure 12-14c, and the magnitude-distance formula, Eq. 9-2a.
15. Polaris is 132 pc from Earth and has an average apparent magnitude of 2.0. What is the star's average luminosity using the magnitude-distance formula (Eq. 9-2a) and Figure 12-14? Based on your answer, is Polaris a type I Cepheid, a type II Cepheid, or a RR Lyrae? Is Polaris a giant or supergiant?
16. If a Cepheid variable has a 2-day period of pulsation and its period increases by 1 second, how late will it be in reaching maximum light after 1 year? (*Hint*: How many cycles will it complete in a year?)

SENSE OF PROPORTION

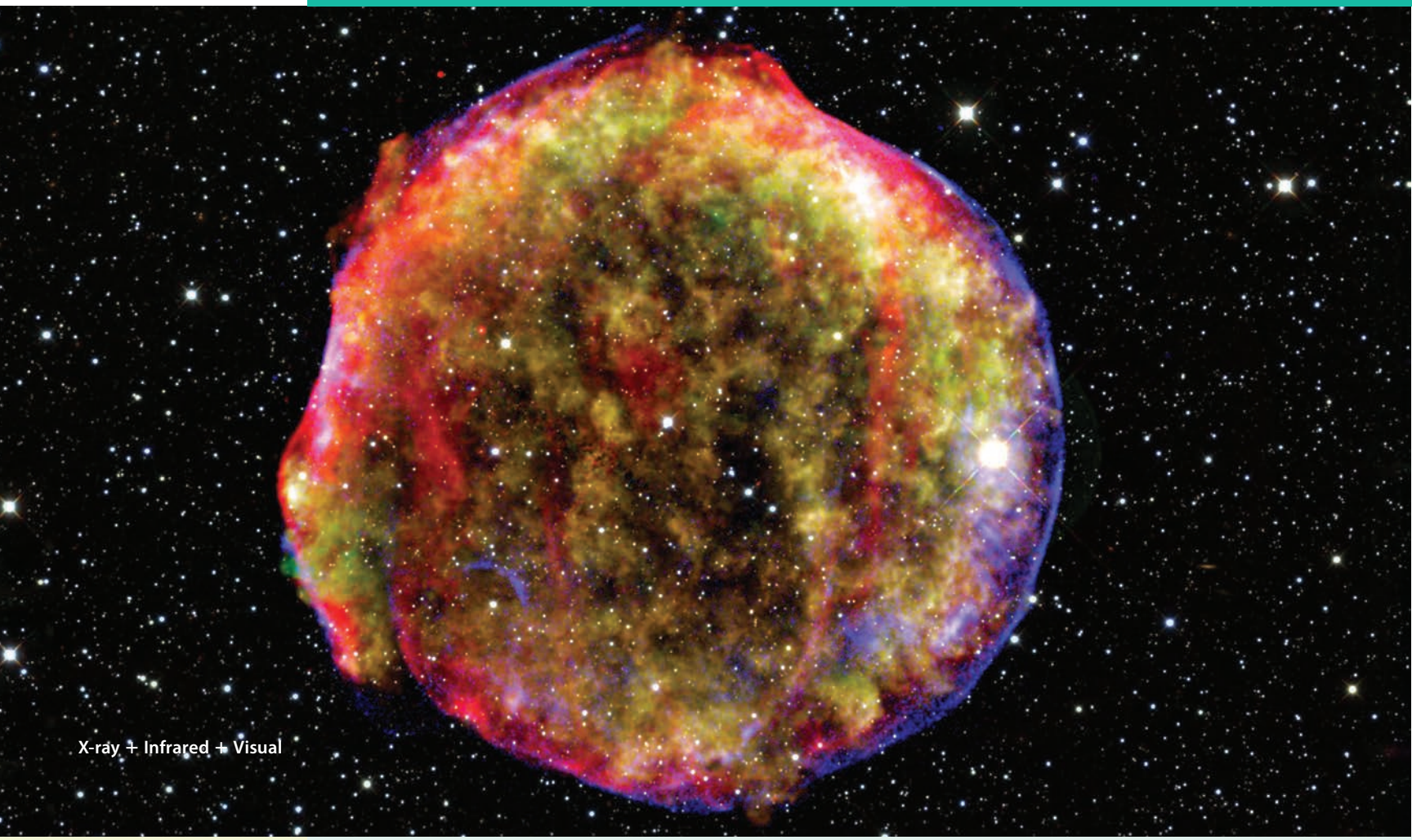
1. The Sun will be on the main sequence fusing hydrogen in its core about 10,000 times longer than an O5 star. If you represent the life of the Sun by one month (30 days), how long would an O5 star remain on the main sequence? Express your answer in minutes.
2. The two components of Eta Carinae have a total luminosity 4 million times that of the Sun. If the Sun is represented in a model by a 40-watt desk light bulb, how many watts would represent the Eta Carinae system? (Note that a city of 100,000 people consumes, on average, roughly 100 million watts of electricity.)

LEARNING TO LOOK

1. Look at the model of the Sun's interior in Figure 12-2. At what radial distance from the center does the summed luminosity reach 100 percent? How much mass and density are at and below that layer? Using these stellar properties and values, explain why the luminosity reaches 100 percent at this radial distance as opposed to at $r = 0\ R_\odot$ or at $r = 1.0\ R_\odot$. Why does the luminosity from this radius to the surface stay at 100 percent?
2. Look at Figure 12-4b. Why did the artist draw alternating patterns of dark and red stripes on the brown dwarf? What is thought to be occurring?
3. In the photograph of the Pleiades in Figure 3 of the Chapter 12 Concept Art, no bright red stars can be seen. Use the H–R diagram to explain why the brightest stars are blue. Were there once bright red stars in this cluster?
4. Look at the photograph of the star cluster M67 in Figure 3 of the Chapter 12 Concept Art. Why are no bright blue stars seen in this cluster?
5. The star cluster in this photo contains many hot, blue, luminous stars. Which kind of cluster is it? Sketch its H–R diagram and discuss its probable age.

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/R. O'Connell (Univ. of Virginia), F. Paresce (NIA), E. Young (USRA/SOFIA)

13 The Deaths of Stars



Courtesy of X-ray: NASA/CXC/SAO, Infrared: NASA/JPL-Caltech; Visual: MPIA, Calar Alto, O.Krause et al.

▲ Remnant cloud from a "nova" (new star), now known to have been a supernova explosion representing the death of a massive star, that was visible to the naked eye in the constellation Cassiopeia from 1572 to 1574.

Perhaps you were surprised to learn in the previous chapters that stars are born and that they grow old. In fact, astronomers can tell the life stories of stars right to their ends. In this chapter you will learn how stars die; as you follow those stories, you see more examples of how scientists test hypotheses against evidence. Here you will find answers to five important questions:

- How will the Sun and other low-mass stars die?
- What happens if an evolving star is a member of a close binary system?
- How do massive stars die?
- What do we learn about stellar evolution from observations of supernovae and supernova remnants?
- What will be the ultimate fate of Earth as the Sun evolves and dies?

Astronomy is interesting because it is about us. As you think about the deaths of stars, you are also thinking about the long-term safety of Earth as a home for life and about the ultimate fate of our Sun, our Earth, and the atoms of which you are made.

> **Natural laws have no pity.**
>
> ROBERT HEINLEIN,
> *THE NOTEBOOKS OF LAZARUS LONG*

GRAVITY IS PATIENT, so patient it can kill stars. Of course, stars are not living things, and they don't really die. Stars "live" by generating tremendous amounts of energy and resisting their own gravity, but they do not have an infinite supply of fuel for nuclear reactions. When their fuel runs out, they can be said to "die."

Stars don't disappear when they die. You will discover that they can produce complex nebulae during their death throes and leave behind some of the strangest and most beautiful objects in the Universe (for example, see the image that opens this chapter).

The mass of a star determines its fate. Lower-mass stars like the Sun die relatively gentle deaths, but the most massive stars explode violently. Also, stars that orbit close to another star can have their evolution modified in complicated ways. To follow the evolution of stars to their graves, you can divide stars into two categories: low-mass (lower-main-sequence) stars, including the Sun, and high-mass (upper-main-sequence) stars.

© David Malin/UK Schmidt Telescope/DSS/AAO

▲ **Figure 13-1** The nearest star to our solar system, Proxima Centauri, is seen against the background of the stars in a dense region of the Milky Way. Proxima, as it is sometimes called, seems bright in this image only because it is so close, only 4.2 ly away. It is a red dwarf of spectral class M5, with a luminosity of only 1/600 $L_\odot$ and an estimated mass of 0.12 $M_\odot$. Proxima can remain a main-sequence star, steadily fusing hydrogen into helium, for trillions of years.

13-1 Low-Mass Stars

The stars of the lower main sequence have relatively low masses, and they face similar fates as they exhaust their nuclear fuels.

As you learned in Section 12-2, when a star exhausts one nuclear fuel, its interior contracts and grows hotter until the next nuclear fuel ignites. The interior of a star heats up as contraction converts gravitational energy into thermal energy. But low-mass stars have relatively low gravitational potential energy, so there is a limit to how hot their interiors can get, and that restricts the fuels they can ignite. The lowest-mass stars, for example, cannot get hot enough even to ignite helium fusion.

You have also learned that structural differences divide the lower-main-sequence, lower-mass stars into two subcategories—very-low-mass stars and medium-mass stars such as the Sun. The critical difference between the two groups is the extent of convection in their interiors (look back to Figure 11-13). If the star is convective throughout its interior so that the hydrogen fuel is constantly mixed, the resulting evolution of the star is strongly affected.

13-1a Red Dwarf Stars

The lowest-mass objects below the end of the main sequence, brown dwarfs, have masses less than 0.08 solar mass ($M_\odot$) and can't get hot enough to ignite hydrogen fusion. These objects just cool down and fade away after they form.

Stars with masses between 0.08 and 0.5 $M_\odot$—red dwarfs, spectral type M—can ignite hydrogen fusion and survive a long time. Their low masses mean that they have very little weight to support. Their pressure–temperature thermostats are set low, and they consume their hydrogen fuel very slowly. Furthermore, unlike medium-mass stars such as the Sun, red dwarf stars are convective throughout their interiors, like pots of soup that are constantly stirred, so all the hydrogen fuel is eventually carried to their cores. The Sun will be able to fuse only about 15 percent of its hydrogen while it is a main-sequence star, but red dwarfs can fuse nearly 100 percent of their hydrogen. If you calculate the life expectancy of one of these stars using the stellar life expectancies equation (Eq. 12-1), you will find that, because of their low masses, red dwarf stars can be expected to live many times longer than the Sun. In fact, models of a 0.12-$M_\odot$ red dwarf such as Proxima Centauri (**Figure 13-1**) predict it will take 4 trillion years (4000 billion years) to use up its hydrogen fuel, almost 400 times as long as the Sun.

Because red dwarf stars use up their hydrogen fuel uniformly throughout their interiors, they never develop a hydrogen-fusion shell, so they never become giant stars. Red dwarfs slowly convert their hydrogen into helium but are not massive enough ever to ignite helium as fuel. When they finally run out of hydrogen they will contract and die slow, unremarkable deaths, going straight from the main sequence to become white dwarf remnants.

You will learn in a later chapter about evidence that the Universe is about 14 billion years old, so no red dwarf should have exhausted its fuel yet. Every red dwarf that has ever been born is still a main-sequence star, still fusing hydrogen into helium and still shining. So then, which stars evolved into the many white dwarfs that are observed today in the Galaxy?

13-1b Medium-Mass Stars

Stars that begin their lives with masses between 0.5 and 8 $M_\odot$, including the Sun, fuse hydrogen, and later, helium, but their cores can never become hot enough to ignite carbon, the next fuel in the sequence (look back to Table 12-3). When medium-mass stars reach that impasse, they can no longer maintain their stability by fusion reactions. The carbon–oxygen core is a dead end for them. (Note that these mass limits are uncertain, as are other limits described in later sections. The evolution of stars is highly complex, and some of the model calculations involved are beyond the capability even of the most advanced software running on supercomputers.)

Although a medium-mass star can't fuse the carbon and oxygen left in its core by earlier helium fusion, it can continue to fuse helium in a shell. As the helium-fusing shell burns outward, it leaves carbon and oxygen "ashes" behind, and that increases the mass of the carbon–oxygen core. Because the core can't get hot enough to fuse carbon, the weight pressing down on it cannot be resisted, so the core contracts. The energy released by the contracting core, plus the energy generated in the helium- and hydrogen-fusion shells, flows outward and makes the envelope of the star expand and cool further.

These changes cause the star to loop back again into the red-giant region of the H–R diagram (Figure 12-11), and this second time around it becomes very large. Its radius may become larger than the radius of Earth's orbit, and its surface may become as cool as 2500 K. And, such a star can lose large amounts of mass from its surface because of a strong wind.

13-1c Mass Loss and Stellar Evolution

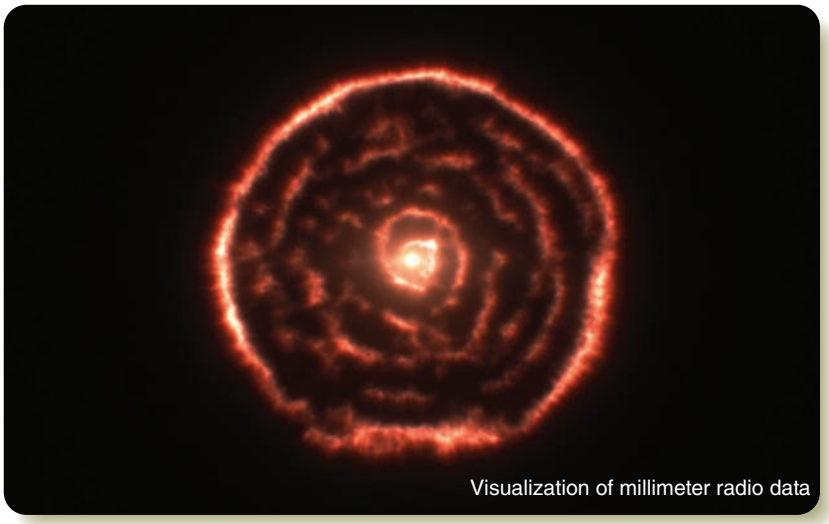
It's not hard to find evidence that stars like the Sun lose mass. Observations show that the Sun itself is losing mass through the solar wind, a comparatively gentle breeze of gas that blows outward from the solar corona and carries mass into space. The Sun loses only about 0.00001 of its mass per billion years. Even over the entire main-sequence lifetime of the Sun, the solar wind will not significantly reduce the Sun's mass.

However, the spectra of some stars contain clear evidence of rapid mass loss. Ultraviolet and X-ray spectra of many stars reveal strong emission lines, and you can use Kirchhoff's laws (look back to Figure 1b in the Chapter 7 Concept Art: Atomic Spectra) to conclude that the star's outer layers must be highly ionized. That means the stars must have hot chromospheres and coronas like the Sun's and therefore stellar winds blowing outward. You can find further evidence in the blueshifted absorption lines observed in the spectra of some stars. The blueshifts are Doppler shifts produced as the gas flowing out of the star moves toward Earth.

Red giant stars lose mass much more rapidly than the Sun. Because these stars do not show telltale X-ray emission, you can assume that they do not have hot coronas like the Sun's, but other processes can drive mass loss (**Figure 13-2**). Giant stars are so large that gravity is weak at their surfaces, and convection in the cool gas can drive shock waves outward that propel mass loss. In addition, some giants are so cool that specks of carbon dust condense in their atmospheres, just as soot can condense in the updraft from a fire. The pressure of the star's radiation can push this dust and any atoms that collide with the dust completely out of the star. Mass loss from giant stars can be so rapid that astronomers refer to it as a **superwind**.

Rapid mass loss can affect stars dramatically. Observations show that some giant stars are losing mass so rapidly that they can shed an entire solar mass in only 100,000 years, a short time in the evolution of a star. This means that a star that began its existence on the main sequence with, say, a mass of 8 $M_\odot$ could become a giant and reduce its mass to only 3 $M_\odot$ in half a million years.

Mass loss confuses the story of stellar evolution. Astronomers would like to be able to say that stars more massive than a certain limit will evolve one way, and stars less massive will evolve another way. But giant stars may lose enough mass to alter their own evolution. As a result, you need to consider both the initial mass a star has on the main sequence (look again at Table 12-3)



ESO/NAOJ/NRAO/ALMA

◀**Figure 13-2** Outflowing gas with an unusual spiral structure is observed around the red giant star R Sculptoris. Data obtained with the ALMA radio telescope array at the wavelength of an emission line of carbon monoxide (0.87 mm) were used to create this visualization. During the red giant stage, stars periodically undergo thermal pulses because of the instability of a helium-burning shell around the core. Those pulses cause temporary large increases in the stellar wind rate. These observations indicate that R Sculptoris had a thermal pulse about 1800 years ago that lasted for 200 years. A companion star orbiting close to R Sculptoris caused the wind to have a spiral structure. (The ALMA observatory is pictured on the opening page of Chapter 6.)

and the mass it retains after later stages of mass loss. The effects of mass loss in stellar evolution are difficult to calculate, so exact mass limits remain quite uncertain.

13-1d Planetary Nebulae

When a medium-mass star like the Sun expands and becomes a red giant for the second time, its atmosphere cools. As it cools, it becomes more opaque, and light has to push against the atmosphere to escape. At that same stage, model calculations predict that the helium-fusion shell will become narrow and unstable, causing it to flare, which also pushes the atmosphere outward. Because of this outward pressure, an aging giant can expel its outer atmosphere in repeated surges to form one of the most beautiful objects in astronomy: a **planetary nebula**, so called because through a small telescope some of them look like the greenish-blue disk of a planet such as Uranus or Neptune. Keep in mind as you read the rest of this section that a planetary nebula has nothing to do with a planet. It is composed of ionized gases expelled by a dying star.

Look through **Formation of Planetary Nebulae and White Dwarfs** and notice four things:

1. You can understand the characteristics of planetary nebulae by using simple observational principles such as Kirchhoff's laws and the Doppler effect.
2. Astronomers have developed a model to explain planetary nebulae. Real nebulae are more complex than the simple model of a slow wind and a fast wind, but the model provides a way to organize the observed phenomena.
3. Oppositely directed jets (much like bipolar flows from protostars) produce some of the asymmetries seen in planetary nebulae.
4. After the star has lost much of its outer layers, the remains of the star contract, and it becomes a white dwarf.

To ionize the gas and light up a planetary nebula, a star must become a white dwarf with a temperature of at least 25,000 K. Mathematical models show that a collapsing star of less than 0.55 $M_\odot$ can take as long as a million years to heat up enough to ionize its nebula; by that time, the expelled gases are long gone. Models of the Sun are not precise enough to indicate how much mass will be left once it ejects its outer layers. If it is left with too little mass, it may heat too slowly to excite the gases it has ejected and make a planetary nebula. Also, some research suggests that to eject a planetary nebula, a star needs a close binary companion to speed up its spin. The Sun, of course, has no close binary companion.

Formation of planetary nebulae is an area of current observational and theoretical research. Some astronomers compute models that show the Sun will definitely light up a planetary nebula. Other astronomers compute models with slightly different initial conditions indicating that the remains of the Sun will heat too slowly to ionize the gases around it. There are no firm conclusions, and the moment of truth lies at least 7 billion years in the future.

Medium-mass stars such as the Sun die by ejecting gas into space and contracting into white dwarfs. Planetary nebulae provide some evidence regarding the deaths of medium-mass stars. Now you can turn your attention to the evidence revealed by white dwarf remnants.

13-1e White Dwarfs

The census of stars (Chapter 9 Concept Art: The Family of Stars) revealed that white dwarfs, although very faint, are fairly common. Now you recognize white dwarfs as the ends of the lives of stars that fused hydrogen and then helium, failed to ignite carbon, drove away their outer layers, and finally collapsed to form remnants of degenerate matter. The billions of white dwarfs in our Galaxy are the remains of medium-mass stars like the Sun.

The first white dwarf discovered was the faint companion to Famous Star Sirius (Alpha Canis Majoris). In that visual binary system, the bright star is designated Sirius A. The white dwarf Sirius B is 8000 times fainter than Sirius A. The orbital motions of the stars (look back to Figure 9-16) reveal that the white dwarf's mass is 0.98 $M_\odot$, and its blue-white color tells you that its surface is hot, about 25,000 K. Although it is very hot, it has a very low luminosity, so it must have a small surface area; in fact, it is slightly smaller than Earth. Dividing its mass by its volume reveals that it is very dense—more than 2×10^6 g/cm^3. On Earth, a teaspoonful of Sirius B material would weigh more than 10 tons (**Figure 13-3**). Basic observations and simple physics lead to the conclusion that white dwarfs are astonishingly dense.

▲ **Figure 13-3** The degenerate matter inside a white dwarf is so dense that a lump the size of a beach ball, transported to Earth, would weigh as much as an ocean liner.

Formation of Planetary Nebulae and White Dwarfs

NASA/ESA/STScI/AURA/NSF/JPL-Caltech

Visual + Infrared

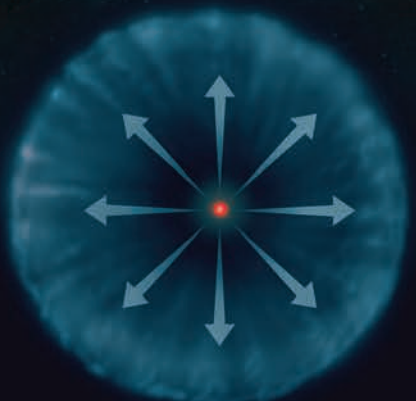
1 Simple observations tell astronomers about the nature of planetary nebulae. Their angular size and distances indicate that their radii range from 0.2 to 3 ly. The presence of emission lines in their spectra implies that they are excited, low-density gas. Doppler shifts show they are expanding at 10 to 20 km/s. If you divide radius by velocity, you find that planetary nebulae are no more than about 100,000 years old. Older nebulae evidently become mixed into the interstellar medium and disappear.

Astronomers find almost 3000 planetary nebulae in the sky. Because planetary nebulae are short-lived formations, you can conclude that they must be a common part of stellar evolution. Medium-mass stars up to a mass of about 8 to 10 $M_\odot$ are destined to die by forming planetary nebulae.

The Helix Nebula is 2.5 ly in diameter, and the radial texture shows how light and winds from the central star are pushing outward.

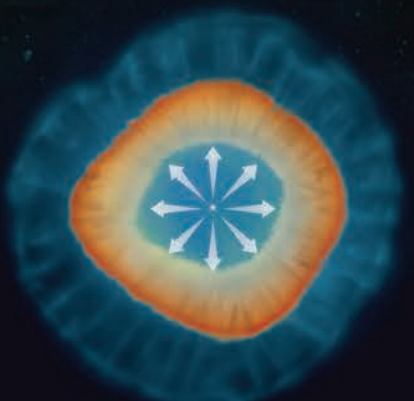
2 The process that produces planetary nebulae involves two stellar winds. First, as an aging giant, the star gradually blows away its outer layers in a slow breeze of low-excitation gas that is not easily visible. Once the hot interior of the star is exposed, it ejects a high-speed wind that overtakes and compresses the gas of the slow wind like a snowplow, while ultraviolet radiation from the hot remains of the central star excites the gases to glow like a giant neon sign.

Slow stellar wind from a red giant



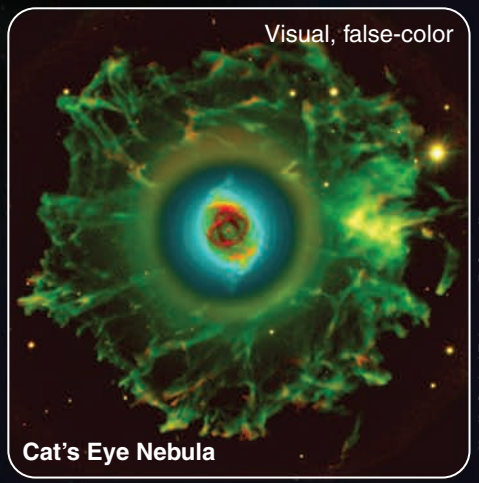
The gases of the slow wind are not easily detectable.

Fast wind from exposed interior



You see a planetary nebula where the fast wind compresses the slow wind.

2a The Cat's Eye, below, lies at the center of an extended nebula that must have been exhaled from the star long before the fast wind began forming the visible planetary nebula. See other images of this nebula on the opposite page.

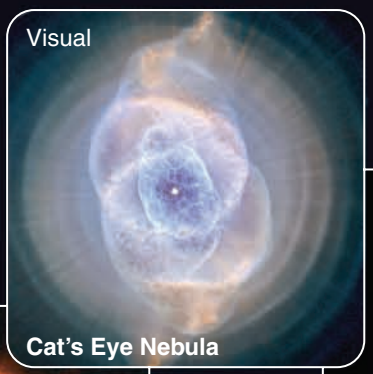


Visual, false-color

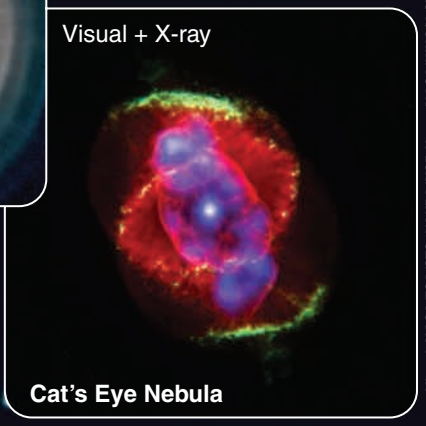
Cat's Eye Nebula

Nordic Optical Telescope/R. Corradi

3 Images from the *Hubble Space Telescope* reveal that asymmetry is the rule in planetary nebulae rather than the exception. A number of causes have been suggested. A disk of gas around a star's equator might form during the slow-wind stage and then deflect the fast wind into oppositely directed flows. Another star or planets orbiting the dying star, rapid rotation, or magnetic fields might cause these peculiar shapes. The Hour Glass Nebula seems to have formed when a fast wind overtook an equatorial disk (*white in the image*). The nebula Menzel 3 shows evidence of multiple ejections, as do many other planetary nebulae.



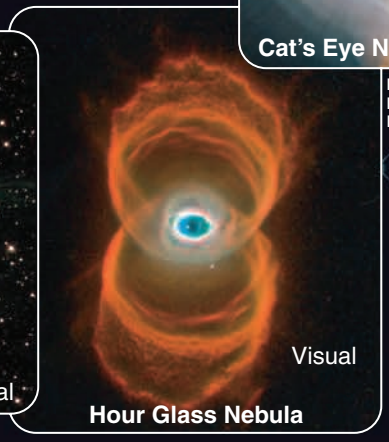
Some shapes suggest bubbles being inflated in the interstellar medium. The Cat's Eye Nebula is shown at left, below, and on the facing page.



Visual: NASA/ESA/STScI/AURA/NSF; X-ray: NASA/UIUC/Y. Chu et al.

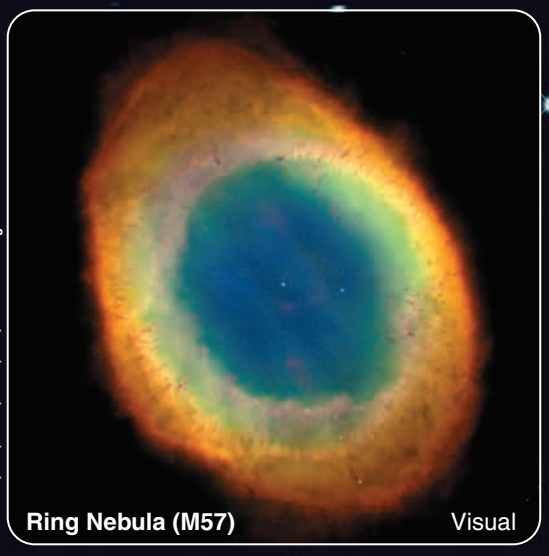
The purple glow in the image above is a region of X-ray-emitting gas with a temperature measured in millions of kelvins. It is apparently driving expansion of the nebula.

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team



NASA/ESA/STScI/AURA/NSF/HEIC/*Hubble* Heritage Team

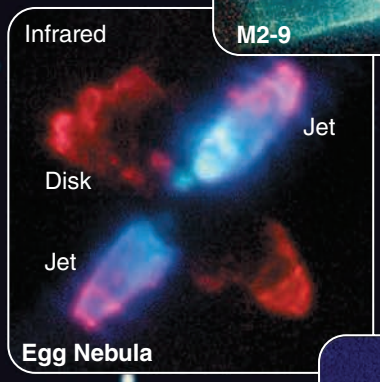
NASA/ESA/STScI/AURA/NSF/WFPC2 Science Team/R. Sahai and J. Trauger (JPL-Caltech)



NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

Some planetary nebulae such as M2-9, at right, are highly elongated. Some astronomers hypothesize that the Ring Nebula, at left, is a tubular shape that happens to be pointed approximately at Earth.

NASA/ESA/STScI/AURA/NSF/B. Balick (Univ. of Washington) and V. Icke (Leiden Univ.)



NASA/ESA/STScI/AURA/NSF/NICMOS IDT/R. Thompson, M. Rieke, G. Schneider, D. Hines (Univ. of Arizona), R. Sahai (JPL-Caltech)

At visual wavelengths, the Egg Nebula is highly elongated, as shown below. The infrared image at left reveals an irregular, thick disk from which jets of gas and dust emerge. Such beams may create many of the asymmetries in planetary nebulae.

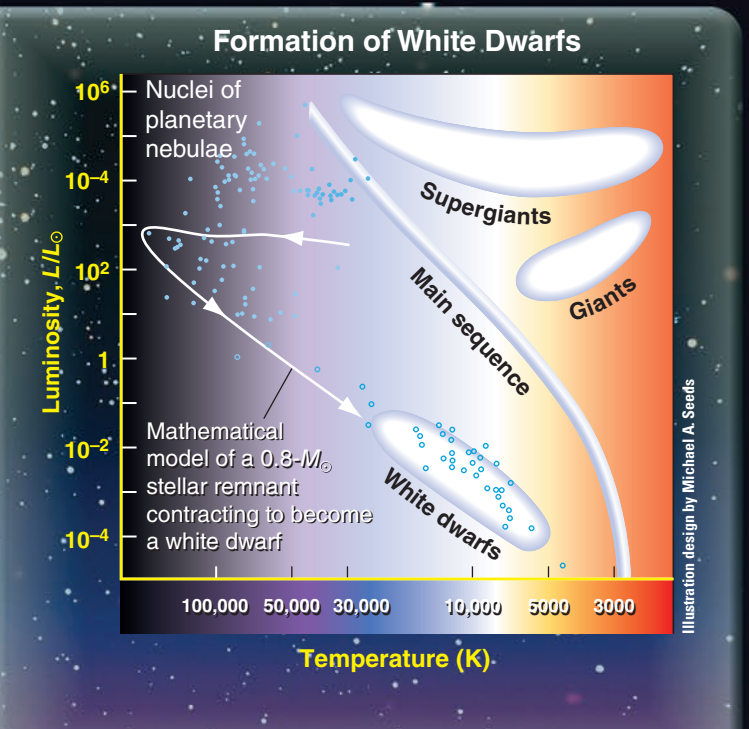
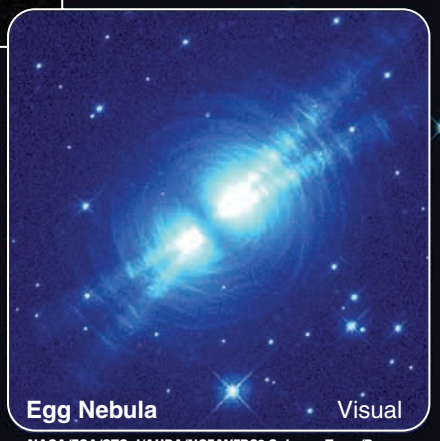


Illustration design by Michael A. Seeds

4 Once an aging giant star expels its surface layers into space to form a planetary nebula, the remaining hot interior collapses into a small, intensely hot object containing a carbon and oxygen interior surrounded by hydrogen and helium fusion shells and a thin atmosphere of hydrogen. The fusion gradually dies out, and the core of the star evolves to the left of the conventional H–R diagram to become the intensely hot nucleus of a planetary nebula. Mathematical models show that these nuclei cool slowly to become white dwarfs.



NASA/ESA/STScI/AURA/NSF/WFPC2 Science Team/R. Sahai and J. Trauger (JPL-Caltech)

Toward Ultimate Causes

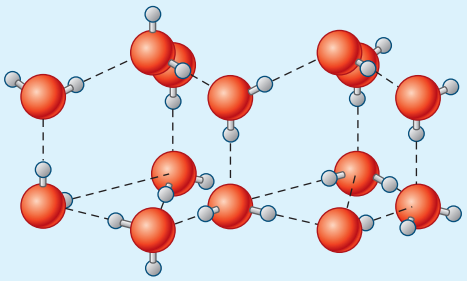
How does a scientist's search for natural causes lead into the world of subatomic particles?

Scientists search for causes. They are not satisfied to know that a certain kind of star dies by exploding. They want to know why it explodes. They want to find the causes for the natural events they see, and that search for ultimate causes often leads into the subatomic world.

For example, why do icebergs float? When water freezes, it becomes less dense than liquid water, so it floats. That answers the question, but you can search for a deeper cause.

Why is frozen water (ice) less dense than liquid water? Water molecules are made up of two hydrogen atoms bonded to an oxygen atom, and the oxygen is so good at attracting electrons that the hydrogen atoms are left needing a bit more negative charge, and they are attracted to atoms in nearby molecules. That means the hydrogen atoms in water are constantly trying to stick to other molecules. When water is warm, the thermal motion prevents these hydrogen bonds from forming, but when water is cold enough, the molecules move slowly and the hydrogen atoms link the water molecules together to form ice. Because of the angles at which the bonds form, open spaces are left between molecules, and that makes ice less dense than water. That's why ice floats.

Scientists can continue this search for causes to smaller and smaller sizes and more and more fundamental questions. Why do electrons have negative charge? What is charge? Nuclear particle physicists are trying to understand those properties of matter. Sometimes the properties of very large things such as supernovae are determined by the properties of the tiniest particles. Science is exciting because the simple observation that ice floats in your lemonade can lead you toward ultimate causes and some of the deepest questions about how nature works.



Ice has a low density and floats because the hydrogen atoms (*gray*) in water molecules form weak bonds (*dashed lines*) to oxygen atoms (*red*) in other molecules when water freezes.

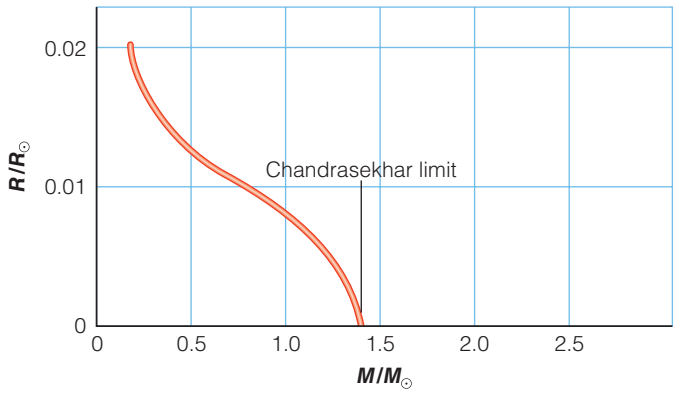
A normal star is supported by energy flowing outward from its core, but a white dwarf cannot generate energy by nuclear fusion. It has exhausted its hydrogen and helium fuels and converted them into carbon and oxygen. When a star collapses into a white dwarf, it converts gravitational energy into thermal energy. Its interior becomes very hot, but it cannot get hot enough to fuse the carbon and oxygen nuclei in its interior. Instead, the star contracts until it becomes degenerate (Section 12-2b, especially Figure 12-8). Although a tremendous amount of energy flows out of its hot interior by conduction through the degenerate matter, that energy flow does not support the star. As you learned in the previous chapter, it is nearly impossible to compress degenerate matter. A white dwarf is supported against its own gravity by the inability of its degenerate electrons to pack into a smaller volume. In this case, the properties of white dwarfs, which are objects as big as Earth, are determined by the properties of subatomic particles. Astronomers often find that the causes of celestial phenomena lead step by step down into the subatomic world (**How Do We Know? 13-1**).

The Sun will leave behind a white dwarf with an interior that is mostly carbon and oxygen nuclei among a whirling storm of degenerate electrons. The degenerate electrons exert the pressure needed to support the star's weight, but most of the star's mass is represented by the carbon and oxygen nuclei. The interior of a white dwarf begins as a degenerate gas, but theory predicts that, as the star grows older and cools, these ions will lock together to form a crystal lattice. So, there is some truth in the image of white dwarfs, at least very old ones, as great crystals of carbon and oxygen. (Note that the Universe is not old enough for this to have happened yet to any white dwarf.)

Near the surface of a white dwarf, where the pressure is lower, a layer of normal (nondegenerate) ionized gases makes up a hot atmosphere The tremendous surface gravity of a white dwarf—100,000 times that of Earth—affects its atmosphere in strange ways. The heavier atoms in the atmosphere tend to sink, leaving the lightest gases at the surface. Astronomers see some white dwarfs with atmospheres of almost pure hydrogen, whereas others have atmospheres of nearly pure helium (gas that was never close enough to the core of the star to be fused into carbon). Still others, for reasons not well understood, have atmospheres that contain traces of heavier atoms. In addition, the powerful surface gravity pulls the white dwarf's atmosphere down into a very shallow layer. If Earth's atmosphere were equally shallow, people on the top floors of skyscrapers would have to wear oxygen masks.

If we define a star to be a sphere of gas generating energy by nuclear fusion, then a white dwarf is not a true star. It generates no nuclear energy, and except for a thin layer at the surface, its matter is almost totally degenerate. Instead of calling a white dwarf a "star," you might refer to it as a **compact object**. In the next chapter, you will meet two other types of compact objects that also are remnants of burned-out stars, neutron stars and black holes.

A white dwarf's future is dim. Degenerate matter is a very good thermal conductor, so heat flows to the surface and escapes into space, and the white dwarf gets fainter and cooler, so the point representing its luminosity and temperature moves downward and to the right in the H–R diagram. As it radiates energy



▲ **Figure 13-4** The more massive a white dwarf, the smaller its radius is. At a mass of 1.4 $M_\odot$, the Chandrasekhar limit, a white dwarf would have a radius of zero. This means that if the contracting remnant of a star after the red giant phase is more massive than 1.4 $M_\odot$, it can't become a white dwarf.

into space, its temperature gradually falls, but it cannot shrink any smaller because its degenerate electrons cannot get closer together. Because the white dwarf contains a tremendous amount of heat, it needs billions of years to radiate that heat through its small surface area. Eventually, such objects are predicted to become cold and dark, so-called **black dwarfs**. Our Galaxy is not old enough to contain black dwarfs. The coolest (and therefore, presumably, the oldest) white dwarfs in our Galaxy have surface temperatures just a bit cooler than the Sun's.

One more surprising and important fact about white dwarfs is predicted by mathematical models. The equations predict that if mass is added to a white dwarf, its radius will *shrink* slightly because added mass increases its gravity and squeezes it tighter. If enough is added to raise its total mass to about 1.4 $M_\odot$, its radius will shrink to zero (**Figure 13-4**). This is called the **Chandrasekhar limit** after Subrahmanyan Chandrasekhar, the astronomer who discovered it. (Subrahmanyan was his family name, Chandrasekhar his given name.) It seems to imply that a star more massive than 1.4 $M_\odot$ cannot become a white dwarf unless it sheds mass in some way.

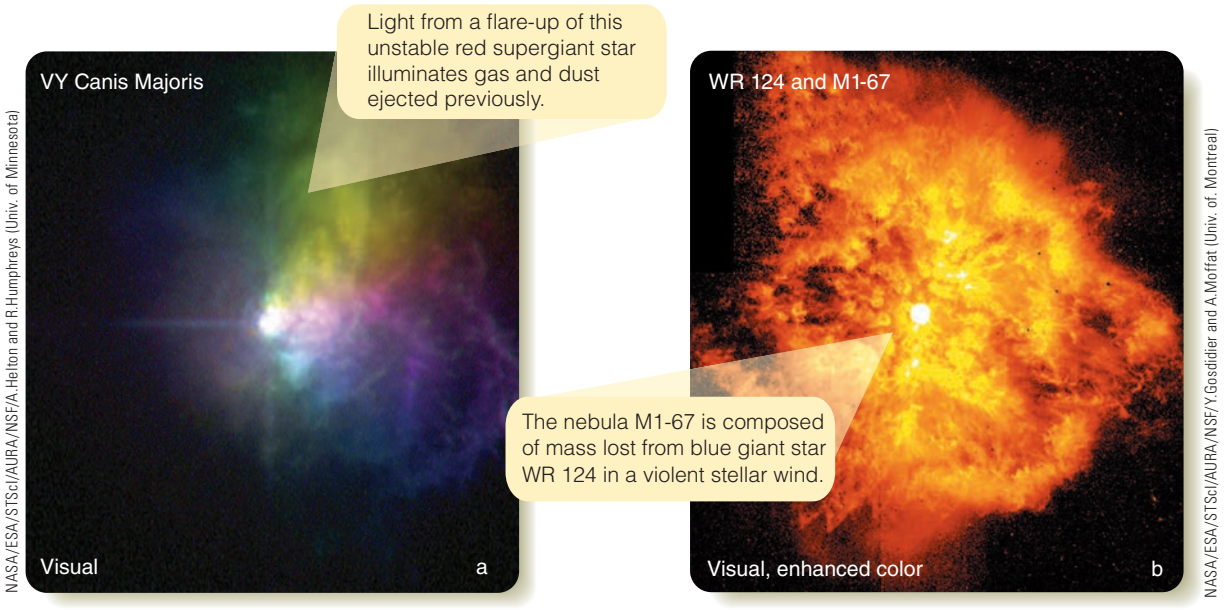
As you learned in this chapter, aging giant stars do lose mass (**Figure 13-5**). This suggests that stars that start their lives more massive than the Chandrasekhar limit can eventually end as white dwarfs if they reduce their mass enough. Model calculations indicate that a star that forms with as much as 8 or even 10 $M_\odot$ probably can reduce its mass to 1.4 $M_\odot$ before it collapses. Consequently, stars with a wide range of "medium" masses eventually become white dwarfs. This explains why white dwarfs are so common.

Practicing SCIENCE

Why do astronomers conclude that large numbers of stars produce planetary nebulae when they die? This argument is based on observations that planetary nebulae typically are only a light-year or so in radius and have Doppler shifts indicating they are expanding at 10 to 20 km/s.

Dividing the radii of planetary nebulae by their expansion velocities tells you that a typical planetary nebula is only about 10,000 years old. This means that the nebulae don't last very long. Nevertheless, astronomers find almost 3000 of them. To be so common but so short lived, planetary nebulae must be produced in large numbers as medium-mass stars blow their outer layers into space.

Now review another conclusion made by astronomers that is based on evidence: ***How do observations of Famous Star Sirius show that white dwarf stars are very dense?***


Light from a flare-up of this unstable red supergiant star illuminates gas and dust ejected previously.
VY Canis Majoris
Visual
a
NASA/ESA/STScI/AURA/NSF/A.Helton and R.Humphreys (Univ. of Minnesota)
The nebula M1-67 is composed of mass lost from blue giant star WR 124 in a violent stellar wind.
WR 124 and M1-67
Visual, enhanced color
b
NASA/ESA/STScI/AURA/NSF/Y.Gosdidier and A.Moffat (Univ. of Montreal)

▲ **Figure 13-5** Stars can lose mass if they are very hot or very luminous. (a) The red supergiant VY Canis Majoris is ejecting loops, arcs, and knots of gas as it ages. (b) A massive hot star such as WR 124 constantly loses mass into space. This star is surrounded by nebula M1-67, which is composed of material it has expelled in the past.

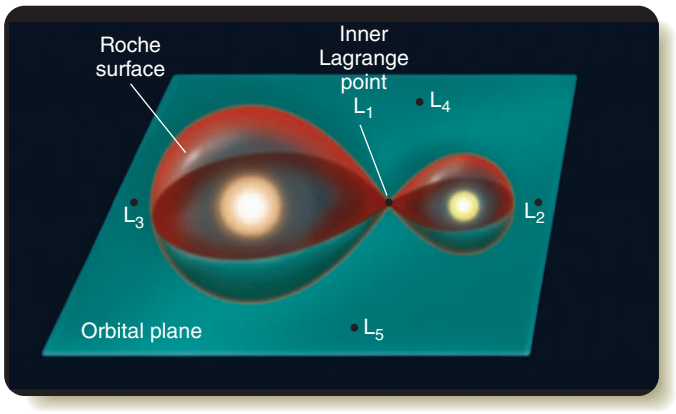
13-2 The Evolution of Binary Stars

So far you have been considering the deaths of stars as if they were all single objects that never interact, but more than half of all stars are members of binary star systems. Most such binaries are far apart, so one of the stars can swell into a giant and eventually collapse without affecting its companion. In some binary systems, however, the two stars orbit close together. When the more massive star begins to expand, it interacts with its companion star in complicated ways. These interacting binary stars are fascinating objects themselves, but they are important because they help explain phenomena such as nova and supernova explosions. In the next chapter, you will see how they can also help astronomers search for black holes.

13-2a Mass Transfer

Binary stars can sometimes interact by transferring mass from one star to the other. The gravitational fields of the two stars, combined with their revolution about the system center of mass, define a dumbbell-shaped volume around the pair of stars called the **Roche lobes**. The gravity of each star controls the motion of matter inside its Roche lobe. The surface of this volume is called the **Roche surface**. The size of the Roche lobes depends on the mass of the stars and on the distance between the stars. If the stars are far apart, the lobes are very large, and the stars easily hang on to their own material. If the stars are close together, however, the lobes are small and can interfere with the evolution of the stars. Matter inside each star's Roche lobe is gravitationally bound to the star, but matter that crosses the Roche surface and leaves a star's Roche lobe can fall into the other star or leave the binary system completely.

The **Lagrange points** are places in the orbital plane of a binary star system where a bit of matter can have extra stability. For astronomers studying binary stars, the most important of these points is the **inner Lagrange (L_1) point**, where the two Roche lobes meet (Figure 13-6). If matter can leave a star and pass the L_1 point, it can easily flow onto the other star. Thus, the L_1 point is the connection through which the stars can transfer matter.



▲ **Figure 13-6** A pair of binary stars controls the region of space located inside the double-lobed Roche surface. The Lagrange points are locations of stability, with the inner Lagrange point (L_1) making a connection through which the two stars can transfer matter.

In general, there are only two ways matter can escape from a star and reach the L_1 point. First, if a star has a strong stellar wind, some of the gas blowing away from it can pass through the L_1 point and be captured by the other star. Second, if an evolving star expands so far that it fills its Roche lobe, which can occur if the stars are close together and the lobes are small, then matter can overflow through the L_1 point onto the other star. Mass transfer driven by a stellar wind tends to be slow, but mass can be transferred rapidly away from an expanding star.

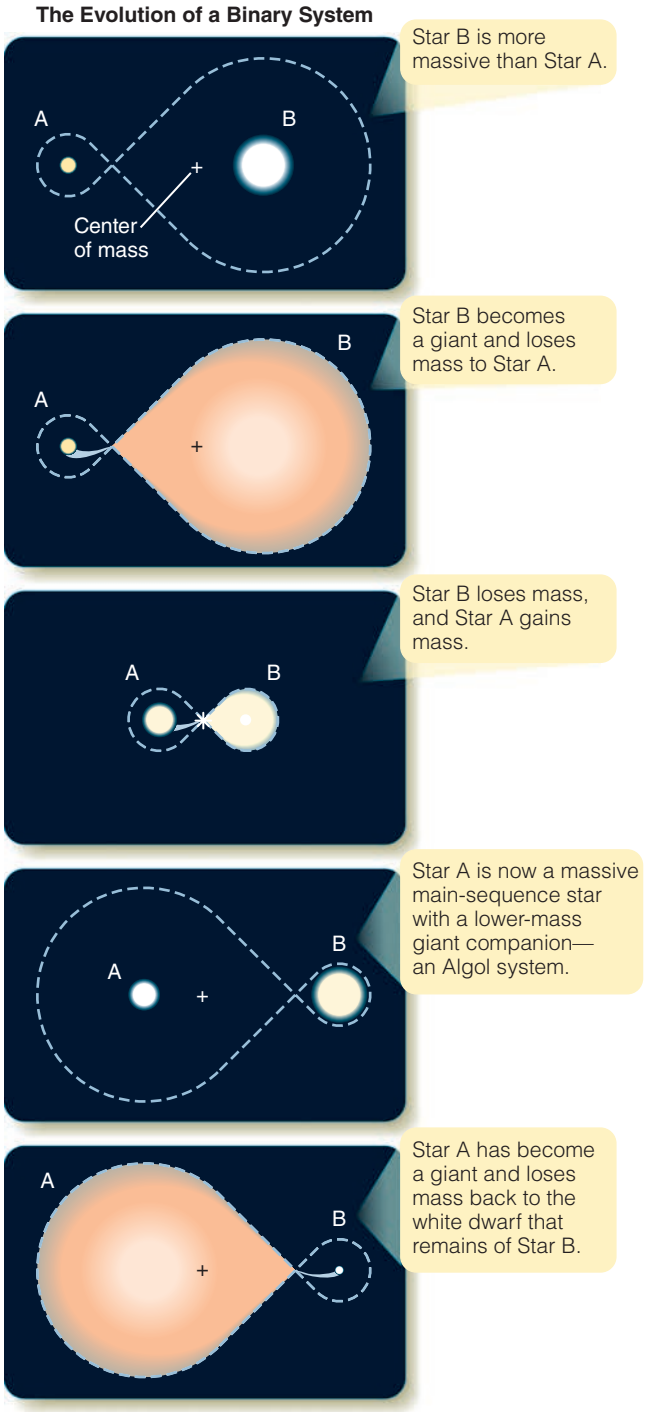
13-2b Mass Transfer and Stellar Evolution

Mass transfer between binary stars provides the answer to a stellar evolution problem that puzzled astronomers for many years. In some binary systems, the less massive star has become a giant, while the more massive star is still on the main sequence. That seems backward. If more massive stars evolve faster than lower-mass stars, how could the low-mass star leave the main sequence first? This is called the Algol paradox, after the binary system Algol (also known as Beta Persei; look back to Figure 9-21).

Mass transfer explains how this could happen. Imagine a close binary system that contains a 5-$M_\odot$ star and a 1-$M_\odot$ companion (Figure 13-7). The two stars formed at the same time, so the more massive star evolves faster and leaves the main sequence first. When it expands into a giant, it fills its Roche surface and transfers matter to the low-mass companion. The massive star shrinks to become a lower-mass giant star, and the companion gains mass to become a more massive main-sequence star. If you observed such a system after the mass transfer ended, you would see a binary system like Algol, which contains a 5-$M_\odot$ main-sequence star and a 1-$M_\odot$ giant.

Another exotic result of the evolution of close binary systems is the possible merging of the stars. Astronomers see many binaries in which both stars have expanded to fill their Roche surfaces and now spill mass out into space. If the stars are close enough together and expand rapidly enough, model calculations indicate it is possible for the two stars to merge into a single, rapidly rotating giant star. Most giants rotate slowly because conservation of angular momentum slowed their rotation as they expanded. Examples of rapidly rotating giant stars are known, however, and they are hypothesized to be the result of merged binary stars. Inside the distended envelope of such a star, the cores of the two stars may even continue to orbit each other until friction slows them down and they sink to the center and merge.

Mass transfer can lead to dramatic violence. The first four frames of Figure 13-7 show how mass transfer could have produced a system like Algol. The last frame shows an additional stage in which the giant star has expelled its outer layers and collapsed to form a white dwarf. The more massive companion has expanded and now transfers matter back onto the white


The Evolution of a Binary System
Star B is more massive than Star A.
A
B
Center of mass
Star B becomes a giant and loses mass to Star A.
Star B loses mass, and Star A gains mass.
Star A is now a massive main-sequence star with a lower-mass giant companion—an Algol system.
Star A has become a giant and loses mass back to the white dwarf that remains of Star B.

▲ **Figure 13-7** A pair of stars orbiting close to each other can exchange mass and modify their evolution.

dwarf. Such systems can become the site of tremendous explosions. To understand how that can happen, you can consider in detail how mass falls into a star.

13-2c Accretion Disks

Matter flowing from one star to another cannot fall directly into the star. Rather, because of conservation of angular momentum, it must flow into a whirling disk around the star.

As you learned in Section 5-2d, rotating objects possess angular momentum. In the absence of external forces, an object maintains (conserves) its total angular momentum. For a common example, consider a bathtub full of water. Gentle currents in the water give it some angular momentum, but you can't see its slow circulation until you pull the stopper. Then, as the water rushes toward the drain, conservation of angular momentum forces it to form an obvious whirlpool.

In a binary star system, mass transferred through the L_1 point toward one of the stars must conserve its angular momentum. If the receiving star is small enough, as in the case of a white dwarf, the incoming material will form a rapidly rotating whirlpool around the star called an **accretion disk** (Figure 13-8).

Two important things happen in an accretion disk. First, the gas in the disk grows very hot as a result of friction and tidal forces. The disk also acts as a brake, shifting angular momentum outward in the disk and allowing the innermost matter to fall into the white dwarf. The interior parts of an accretion disk around white dwarfs and other compact objects are violent, hot places. The temperature of the gas can exceed 1 million kelvins, causing the accretion disk to emit X-rays.

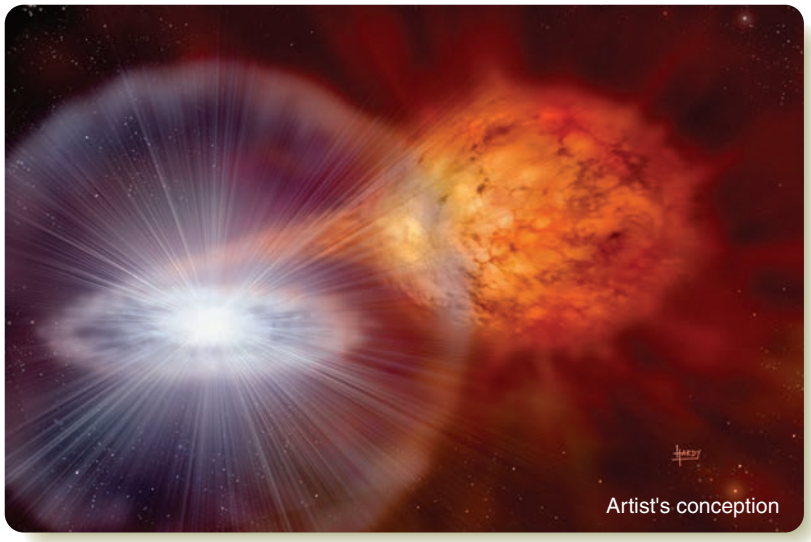
13-2d Nova Explosions

Astronomers occasionally see a **nova**, what seems to be a new star that appears in the sky, grows brighter, and then fades away after a few weeks (Figure 13-9). In fact, *nova* is the Latin word for *new*. Modern astronomers know that a nova is not a new star but an old star flaring up. After a nova fades, astronomers can photograph the spectrum of the remaining faint point of light. Always, they find a short-period spectroscopic binary containing a normal star and a white dwarf. A nova is evidently an explosion involving a white dwarf.

Observational evidence reveals the sequence of events during a nova explosion. As the explosion begins, spectra show blueshifted absorption lines, which tells you the gas is dense and coming toward you at a few thousand kilometers per second. After a few days, the spectral lines change to emission lines, so you know the gas has thinned enough to become transparent. The blueshifts remain, so you can conclude that a spherical shell of gas is continuing to expand into space.

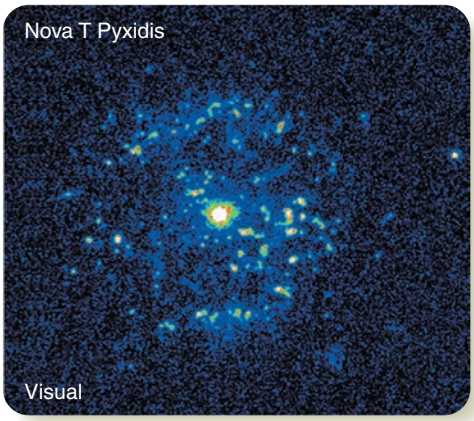
Nova explosions occur in binary systems when mass transfers from a normal star through the L_1 point into an accretion disk around a white dwarf. As the matter loses its angular momentum in the accretion disk, it spirals inward and eventually settles onto the surface of the white dwarf. Because the matter came from the surface of a normal star, it is mostly hydrogen. As it accumulates on the surface of the white dwarf, it forms a layer of ready nuclear fuel.

As the layer of fuel grows deeper, it becomes hotter and denser. Compressed by the white dwarf's gravity, the gas becomes degenerate, much like the core of a Sun-like star approaching the helium flash. In such a gas, the pressure–temperature thermostat



David A. Hardy, www.astroart.org.

▲ **Figure 13-8** Matter from an evolving red giant falls into a white dwarf and forms a whirling accretion disk. Friction and tidal forces can make the disk very hot. Such systems can lead to nova explosions on the surface of the white dwarf, as shown in this artist's conception.



NASA/ESA/STScI/AURA/NSF/M. Shara and R. Williams

▲ **Figure 13-10** Nova T Pyxidis erupts about every two decades, expelling shells of gas into space that this *Hubble Space Telescope* image shows in detail. The shell consists of knots of excited gas that presumably form when a new shell collides with a shell from a previous eruption. Note that at the center of the shell is the system pictured in the Figure 13-8 artist's conception, unresolved in this view.

does not work, so the layer of unfused hydrogen is a thermonuclear bomb waiting to explode. By the time the white dwarf has accumulated between 1 and 50 Earth masses of hydrogen on its surface, the temperature at the base of the hydrogen layer reaches millions of degrees, and the density is 10,000 times the density of water. Suddenly, the nuclear reactions of the proton–proton chain begin to fuse the hydrogen, and the energy released drives the temperature so high that the CNO cycle reactions can operate (Figure 11-12). With no pressure–temperature thermostat to control the fusion, the temperature shoots to 100 million K in seconds. By this time, enough energy has been released to blow the surface layers of the white dwarf into space as a rapidly expanding shell of hot gas, visible from Earth as a nova.

The explosion of its surface layers hardly disturbs the white dwarf or its companion star. Mass transfer quickly resumes, and a new layer of hydrogen fuel begins to accumulate. Thus, you can expect a nova to repeat as often as a hydrogen layer massive enough to explode piles up. Some novae probably take thousands of years to build an explosive layer, but others are observed to take only a few years (**Figure 13-10**).

UC Regents

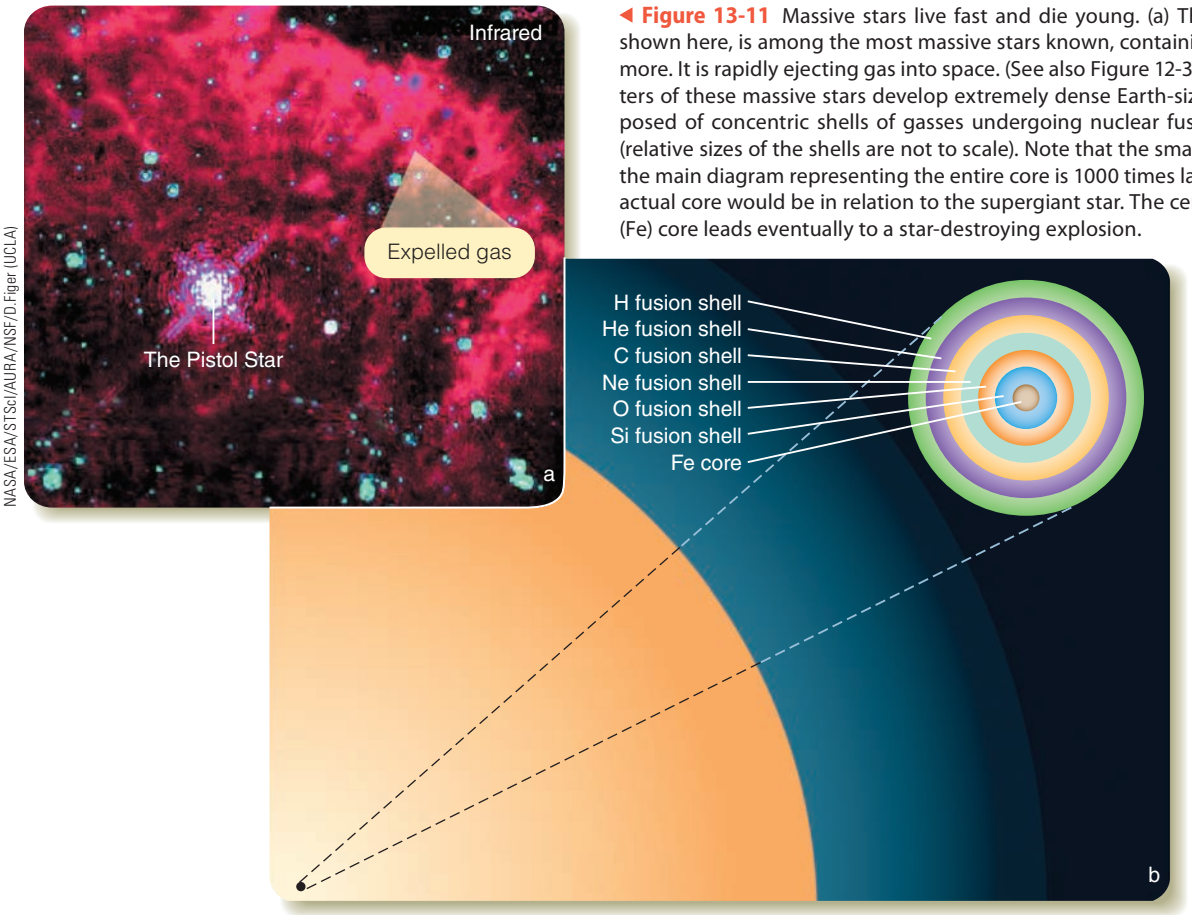
▲ **Figure 13-9** Nova Cygni 1975, photographed near its maximum apparent brightness of about 2nd magnitude and later when it had declined to about 11th magnitude (by a factor of 4000).

Practicing **SCIENCE**

What is the evidence that a nova is actually an explosion?
To answer this question, a scientist would rely on knowledge of basic spectroscopy.

As soon as a nova is seen, astronomers rush to telescopes to record spectra, and they always see blueshifted absorption lines. The blueshifts are Doppler shifts showing that the near side of the object is coming toward Earth at thousands of kilometers per second. Those absorption lines must be formed by opaque gas seen through thinner gas, much like the atmosphere of a star, so the thin gas must be expanding rapidly outward. Later the spectrum becomes an emission spectrum that is still blueshifted. Kirchhoff's laws tell you that an emission spectrum means the gas must have thinned and become transparent. The blueshift shows that the expansion is continuing.

Continue your work as a scientist studying stellar evolution. ***How do observations of novae after they have faded provide evidence that white dwarf stars are involved?***



NASA/ESA/STScI/AURA/NSF/D.Figer (UCLA)

◀ **Figure 13-11** Massive stars live fast and die young. (a) The Pistol Star, shown here, is among the most massive stars known, containing 100 $M_\odot$ or more. It is rapidly ejecting gas into space. (See also Figure 12-3.) (b) The centers of these massive stars develop extremely dense Earth-size cores composed of concentric shells of gasses undergoing nuclear fusion reactions (relative sizes of the shells are not to scale). Note that the small black dot in the main diagram representing the entire core is 1000 times larger than the actual core would be in relation to the supergiant star. The centermost iron (Fe) core leads eventually to a star-destroying explosion.

13-3 High-Mass Stars

You have seen that low- and medium-mass stars die relatively quietly as they exhaust their hydrogen and helium and then eject their surface layers, in some cases, forming planetary nebulae. In contrast, massive stars live spectacular lives and then die in violent explosions.

13-3a Nuclear Fusion in Massive Stars

Stars on the upper main sequence have too much mass to end as stable white dwarfs, but their evolution begins much like that of their lower-mass siblings. They consume the hydrogen in their cores and ignite hydrogen-fusion shells. As a result, they expand into red giants or, for the most massive stars, supergiants. Next, their cores contract and fuse helium: first in the core, and then in a shell, producing a carbon–oxygen core. So far, this sequence of stages is the same for high-mass stars as for medium-mass stars like the Sun.

Then, although a massive star can lose significant mass as it ages, if it still has a mass greater than about 4 $M_\odot$ when its carbon–oxygen core contracts, it can ignite carbon fusion at a temperature of about 1 billion kelvins. Carbon fusion produces more oxygen and neon. As soon as the carbon is exhausted in the core, the core contracts, and carbon ignites in a shell. This pattern of core ignition of a particular nuclear fuel followed by shell ignition of that fuel continues with fuel after fuel, larger and larger nuclei. The star develops a layered structure shown in **Figure 13-11**, with a hydrogen-fusion shell above a helium-fusion shell above a carbon-fusion shell, and so on. Carbon fuses to make oxygen, neon, sodium, and magnesium; oxygen, neon, sodium and magnesium fuse to make silicon, sulfur, and phosphorus; and finally silicon fuses to make iron (Table 12-3).

The fusion of successive nuclear fuels goes faster and faster as the massive star evolves (**Table 13-1**). Recall that massive stars must consume their fuels rapidly to support their great weight, but other factors also cause the heavier fuels like carbon,

Table 13-1 Heavy-Element Fusion in a 25-$M_\odot$ Star

Fuel	Time	Percentage of Lifetime
H	3,000,000 years	88
He	400,000 years	12
C	600 years	0.018
O	0.5 years	0.000015
Si	1 day	0.000000081

oxygen, and silicon to fuse at increasing speeds. For one thing, the amount of energy released per fusion reaction decreases as the mass of the fusing atom increases. To support its weight, a star must fuse oxygen much faster than it fused hydrogen. Also, there are fewer nuclei in the core of the star by the time heavy nuclei begin to fuse. Four hydrogen atoms make a helium nucleus, and three helium atoms make a carbon, so there are 12 times fewer nuclei of carbon available for fusion than there were of hydrogen.

Another important factor in the pacing of massive star evolution is the production of neutrino–antineutrino particle pairs in the star's hot core. Because these particles almost never interact with normal matter (Section 8-3b) they race out of the star carrying away some of the energy, speeding the contraction of the core. This is yet another reason why the fusion of heavy elements goes very quickly in massive stars. Hydrogen core fusion can last 3 million years in a 25-$M_\odot$ star, but that same star will fuse the oxygen in its core in six months and its silicon in a day.

13-4 Supernova Explosions

Heavy-element fusion ends with iron because nuclear reactions that use iron as a fuel cannot produce energy. Nuclear reactions can produce energy if they proceed from less tightly bound nuclei to more tightly bound nuclei. As indicated by the curve of binding energy shown in Figure 8-15, both nuclear fission and nuclear fusion produce nuclei that are more tightly bound than the starting fuel. Notice that iron is the most tightly bound nucleus of all. No nuclear reaction, fission or fusion, that starts with iron can produce a more tightly bound nucleus, and this means that iron is a dead end for nuclear reactions.

13-4a An Iron Core: Looming Catastrophe

When a massive star develops an iron core, nuclear fusion cannot produce energy, and the core contracts and grows hotter. The shells around the core burn outward, fusing lighter elements into heavier elements and leaving behind more iron, which further increases the mass of the core. When the mass of the iron core exceeds the Chandrasekhar limit of about 1.4 $M_\odot$, it must collapse.

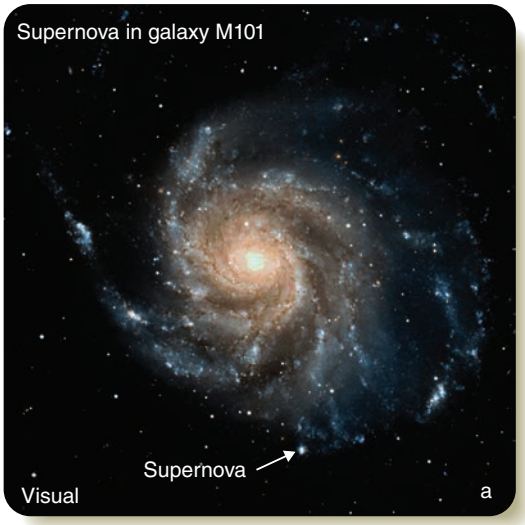
As the core begins to collapse, two processes make it contract even faster. Heavy nuclei in the core capture high-energy electrons, removing thermal energy from the gas. This robs the gas of some of the pressure it needs to support the crushing weight of the outer layers. Also, temperatures are so high that all the photons are gamma-rays, and some carry enough energy to break more massive nuclei into less massive nuclei. In the process, the gamma-rays are absorbed, which robs the gas of energy and allows the core to collapse even faster.

Although a massive star may live for millions of years, its iron core—about 500 km in diameter—collapses in just a few thousandths of a second, triggering a star-destroying explosion.

13-4b Observations of Supernovae

In the 1930s, astronomers realized that some of the novae that had been seen in the sky were much more luminous and longer lasting than the others and dubbed such objects **supernovae** (singular, *supernova*). A few novae detectable with small telescopes appear in the sky each year, but supernovae are so rare that only one or two happen each century in our Galaxy (Figure 13-12a). Now, astronomers understand that a supernova is caused by the violent, explosive death of a star.

The first sign of a supernova is the brightening of the star as its outer layers are blasted outward. As months pass, the cloud of gas expands, thins, and begins to fade. The way it fades can tell astronomers about the death throes of the star. Essentially all of the iron in the core of the star is destroyed when the core collapses, but the violence in the outer layers during the explosion can produce densities and temperatures high enough to trigger short-lived nuclear fusion reactions that produce as much as 0.5 $M_\odot$



NOAO/AURA/NSF/G. Jacoby, B. Bohannan & M. Hanna

NASA/ESA/CXC/SAO/STScI/AURA/NSF/*Hubble* Heritage Team/J. Hughes (Rutgers University)

◀**Figure 13-12** (a) Supernova explosions are rare in any one galaxy, but each year, astronomers see many erupting in other galaxies. (b) Astronomers have discovered many supernovae remnants in our own Galaxy, expanding shells filled with hot, low-density gas produced by past supernova explosions.

of radioactive nickel-56. The nickel gradually decays to form radioactive cobalt, which decays to form nonradioactive (stable) iron. The rate at which the supernova fades matches the rate at which these radioactive elements decay, showing that energy from these decays helps illuminate the expanding gases. Note also that destruction of iron during the core collapse is compensated by production of iron by nuclear reactions in the expanding outer layers.

The fact that nuclear fusion occurs in the outer layers of supernovae testifies to the violence of the explosion. A typical supernova is equivalent to the explosion of more than 10^{28} megatons of TNT, which is about 10 million solar masses of high explosive. And that doesn't count the energy released in a burst of neutrinos that is estimated to be about 100 times larger. (Of course, supernova explosions occur in silence. Science fiction movies and television have led to the **Common Misconception** that explosions in space are accompanied by deep, loud booms. But space is nearly a vacuum, so there can be no sound. Some of the most violent events in the Universe make no noise at all.)

Astronomers know that supernovae occur because they occasionally are seen to flare in other galaxies and because telescopes reveal **supernova remnants**, clouds of debris expanding away from these titanic explosions (**Figure 13-12b**). Modern theory predicts that the collapse of the core in a massive star can eject the outer layers of the star to produce one of the most common kinds of supernova explosions. You will learn about another kind of supernova explosion later in this section.

13-4c Models of Supernovae

A supernova explosion is rapid, violent, and rare. It is an event that is extremely difficult to study directly, which is why astronomers have used mathematical techniques and high-speed supercomputers to model the inside of a star as it explodes. Such models allow astronomers, in a sense, to experiment on supernovae as if the stars were in a laboratory beaker.

Those models reveal that the key to a supernova explosion is the collapse of the iron core. This collapse allows the rest of the star's interior to fall inward, creating a tremendous "traffic jam" as all the nuclei fall toward the center. It is as if all the residents of a state suddenly tried to drive their cars as fast as possible into the center of the capital city. There would be a tremendous traffic jam downtown, and as more cars rushed in, the traffic jam would spread outward into the suburbs. Similarly, as the inner core of the star falls inward, a shock wave (traffic jam) develops and begins to move outward. Such a shock wave was first thought to be the cause of the supernova explosion. Computer models revealed, however, that the shock wave spreading outward through the collapsing star stalls within a few thousandths of a second. Matter falling inward smothers the shock wave and pushes it back into the star. Those computer models predict that the star shouldn't explode.

Then what does cause a supernova? As noted earlier, theory predicts that 99 percent of the energy released as the core collapses appears in the form of neutrinos. In the Sun, neutrinos zip outward unimpeded by the gas of the Sun's interior, but in a supernova collapse, the gas is trillions of times denser than the gas in the Sun's core, nearly as dense as an atomic nucleus. At that density the gas is nearly opaque to neutrinos, so the neutrinos are partially absorbed by the gas. Not only does the tremendous burst of neutrinos remove energy from the core and allow it to collapse even faster, but when the neutrinos are absorbed outside the core they push those layers outward.

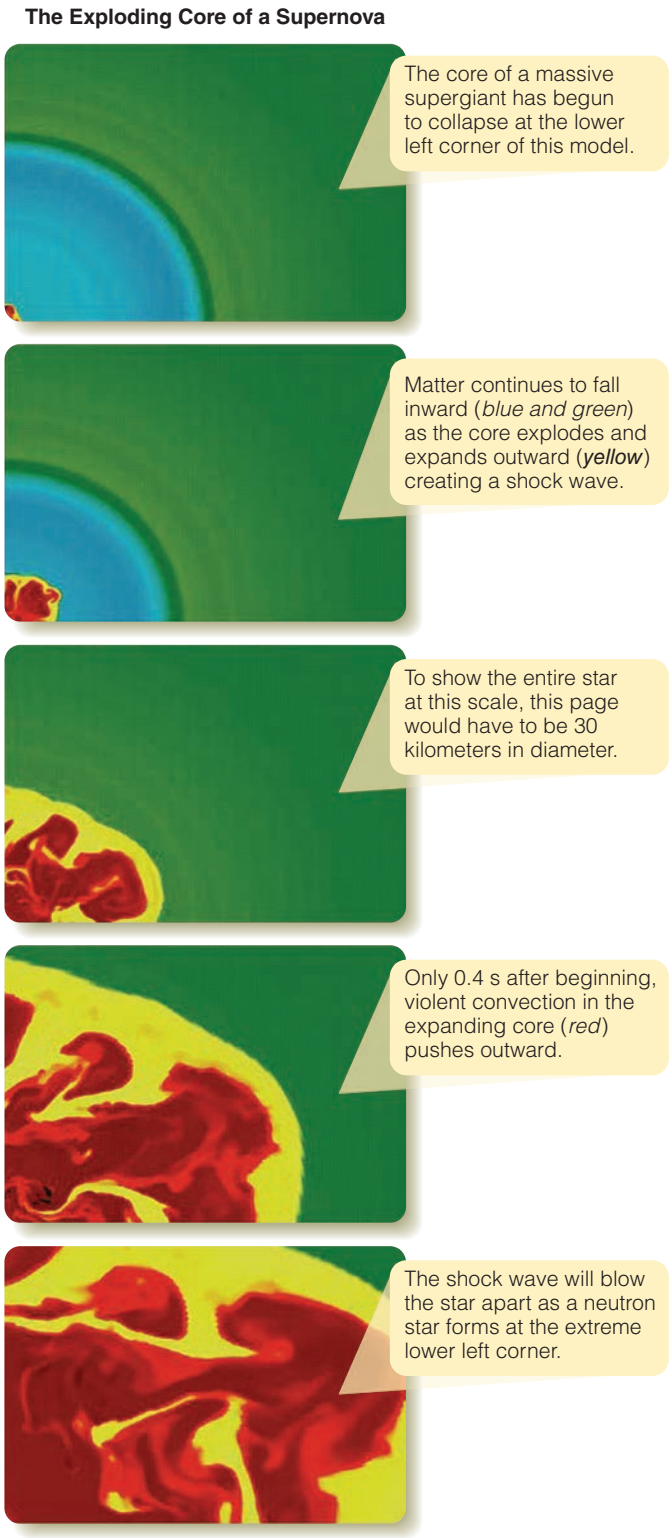
Other processes also boost the explosion. For example, turbulent convection seems to be important. When the collapse begins, the very center of the star forms a highly dense core, which is the beginnings of a neutron star, and the infalling material bounces off that core. As the temperature shoots up, the bouncing material produces highly turbulent convection currents that give the stalled shock wave an outward boost (**Figure 13-13**). The supercomputer models also indicate that magnetic fields and the rotation of the star's core are involved in keeping the shock wave from stalling. Within a second or so, the shock wave begins to push outward again, and after just a few hours it bursts out through the surface, blasting the star apart in a supernova explosion and producing a months-long flare of light that can be seen from billions of light-years away.

13-4d Types of Supernovae

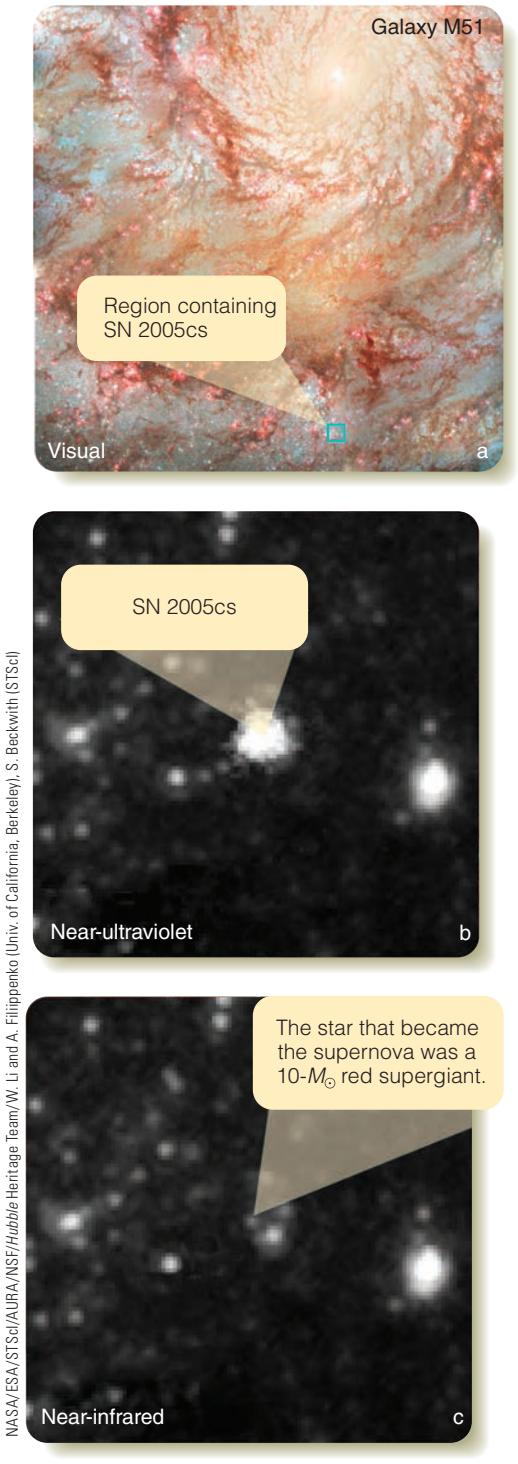
Supernovae are rare, and only a few have been seen in our Galaxy, but astronomers have been able to observe supernovae occurring in other galaxies (**Figure 13-14**; also see Figure 13-12a; note that supernovae are named with the letters SN followed by the year in which they occurred and then a letter code). Collapsing cores of massive stars trigger one kind of supernova explosion, but there are other kinds. From data accumulated over decades, astronomers have noticed that there are two main types. **Type I supernovae** have spectra that contain no hydrogen lines. They reach a maximum brightness about 4 billion times the Sun's luminosity and decline rapidly at first, then more slowly. **Type II supernovae** have spectra that do contain hydrogen lines. They reach a maximum brightness up to about 0.6 billion $L_\odot$ (substantially less than the luminosity of type I supernovae), decline to a temporary standstill, and then fade rapidly. The light curves in **Figure 13-15** summarize the behaviors of these two kinds of supernovae.

The evidence is clear that a type II supernova is the type that occurs when a massive star develops an iron core and collapses. Such supernovae occur in or near regions of active star formation where you would expect to find massive stars. Also, the spectra of type II supernovae contain hydrogen lines, as you would expect from the explosion of a massive star that contains large amounts of hydrogen in its outer layers.

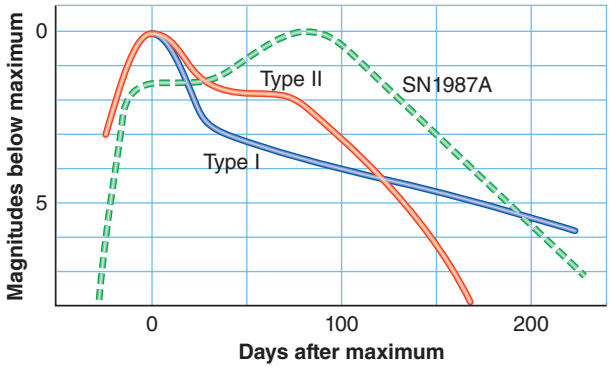
Type I supernovae show no hydrogen in their spectra, which means they can't be caused by the deaths of typical massive stars. In fact, there are two kinds of type I supernovae, and they have dramatically different causes. Both involve binary stars.


The Exploding Core of a Supernova
The core of a massive supergiant has begun to collapse at the lower left corner of this model.
Matter continues to fall inward (*blue and green*) as the core explodes and expands outward (*yellow*) creating a shock wave.
To show the entire star at this scale, this page would have to be 30 kilometers in diameter.
Only 0.4 s after beginning, violent convection in the expanding core (*red*) pushes outward.
The shock wave will blow the star apart as a neutron star forms at the extreme lower left corner.
Courtesy Adam Burrows, John Hayes and Bruce Fryxell

▲ **Figure 13-13** As the iron core of a massive star begins to collapse, intensely hot gas triggers violent convection. Even as the outer parts of the core continue to fall inward, the turbulence blasts outward and reaches the surface of the star within hours, creating a supernova eruption. This diagram is based on mathematical models and shows only the exploding core of the star.



▲ **Figure 13-14** (a) Robot telescopes search every night for supernovae flaring in other galaxies such as Messier 51 (M51). (b) When a supernova such as SN 2005cs is identified, astronomers can record the supernova's rise and fall in brightness (light curve) and obtain spectra to study the physics of exploding stars. (c) If a supernova is seen in a close enough galaxy, it can be possible to identify the pre-explosion star in earlier images.



▲ **Figure 13-15** Type I supernovae decline rapidly at first and then more slowly, but type II supernovae pause for about 100 days before beginning a steep decline. Supernova 1987A was odd in that it did not rise directly to maximum brightness. These light curves have been adjusted to have the same maximum brightness. Generally, type II supernovae are about two magnitudes (a factor of 6) fainter than type I.

A **type Ia supernova** occurs when a white dwarf exceeds the Chandrasekhar limit and begins to collapse. That could happen if a white dwarf gained mass from its companion in a binary star system. It could also happen if two white dwarfs in a binary merged to form a single white dwarf that exceeded the Chandrasekhar limit.

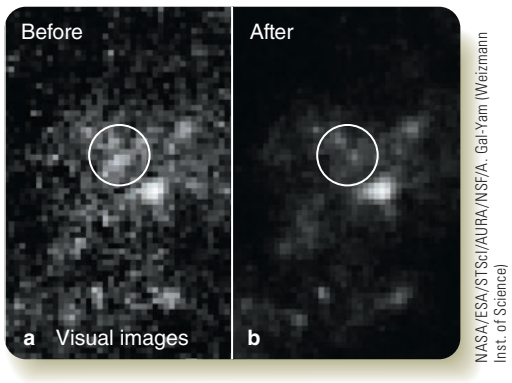
The collapse of a white dwarf is different from the collapse of a massive star because the white dwarf contains usable nuclear fuel—mainly carbon and oxygen. As the collapse begins, the temperature shoots up, but the gas cannot halt the collapse because it is degenerate, so the pressure–temperature thermostat is not working. Even as carbon fusion begins, the increased temperature cannot increase the pressure and make the gas expand and slow the reactions. The lack of a thermostat makes the collapsing white dwarf into a bomb. The carbon–oxygen core fuses suddenly and completely in violent nuclear reactions called **carbon deflagration**. The word *deflagration* means to be totally destroyed by fire, in this case by nuclear fusion. In a flicker of a stellar lifetime, the white dwarf is destroyed as nuclear reactions fuse its carbon–oxygen core into heavy nuclei and blast the outer layers away in a violent explosion that at its brightest is three to six times more luminous than a type II supernova. Nothing remains of the star but an expanding cloud of hot gas. The reason you see no hydrogen lines in the spectrum of a type Ia supernova is that white dwarfs contain at most only a little hydrogen in their outer layers.

You might wonder how an accreting white dwarf can gain mass if the nova explosions blast the accumulated material away. Mathematical models show that nova eruptions do not necessarily blow away all of the mass a white dwarf has gained since the last eruption. Thus, it is possible for a white dwarf to erupt a number of times as a nova, gaining mass all the while, eventually becoming so massive that it collapses in a type Ia supernova explosion. Also, if mass transfer is fast enough, the hydrogen and helium fuse immediately when the matter hits the surface of the white dwarf, so the accumulating layer contains carbon that can't erupt to produce a nova. In either case, the accumulating matter eventually pushes the white dwarf over the Chandrasekhar mass limit, and it collapses to produce a type Ia supernova.

The less common **type Ib supernovae** are thought to occur when a massive star loses its hydrogen-rich outer layers. This can happen to extremely massive stars because their superwinds expel their hydrogen-rich atmospheres. It can also happen if a massive star in a binary system loses its outer layers to its binary companion. The remains of the massive star continue to evolve, develop an iron core, and collapse, producing a supernova explosion that lacks hydrogen lines in the spectrum. Some astronomers have referred to these as "peeled" supernovae because the massive stars have had their hydrogen-rich outer layers peeled away by binary companions.

To summarize, a type II supernova is caused by the collapse of a massive star's iron core, a type Ia is caused by the collapse of a white dwarf, and a type Ib is caused by the collapse of a massive star that has lost its outer envelope of hydrogen. The distinction is evident even in the location of these supernovae. Type Ib and type II supernovae are observed near star-forming regions because they are caused by the collapse of massive stars with such short lives that they can't travel far from their birthplaces. In contrast, type Ia supernovae are caused by the collapse of white dwarfs, but it takes a long time for a medium-mass star to become a white dwarf. That's why type Ia supernovae usually are seen in regions without active star formation.

There may be other types of supernovae caused by other processes that can make a massive star explode. For example, SN 2005gl has been identified with a very luminous, hot supergiant, and astronomers did not expect a star of that type to explode (**Figure 13-16**). That supernova remains a puzzle. Supernovae are so rare that Earth's astronomers may not have seen every type that can occur. Some of the supernovae and supernova remnants that have been studied in detail are described in the next section.



▲ **Figure 13-16** SN 2005gl exploded in the galaxy NGC 266, 200 million ly from Earth. (a) In this highly magnified image made in 1997, the star is visible as a luminous blue supergiant, a class of star not expected to explode as a supernova. (b) In a 2007 image recorded after the explosion faded away, the star is gone.

13-4e Historical Supernovae

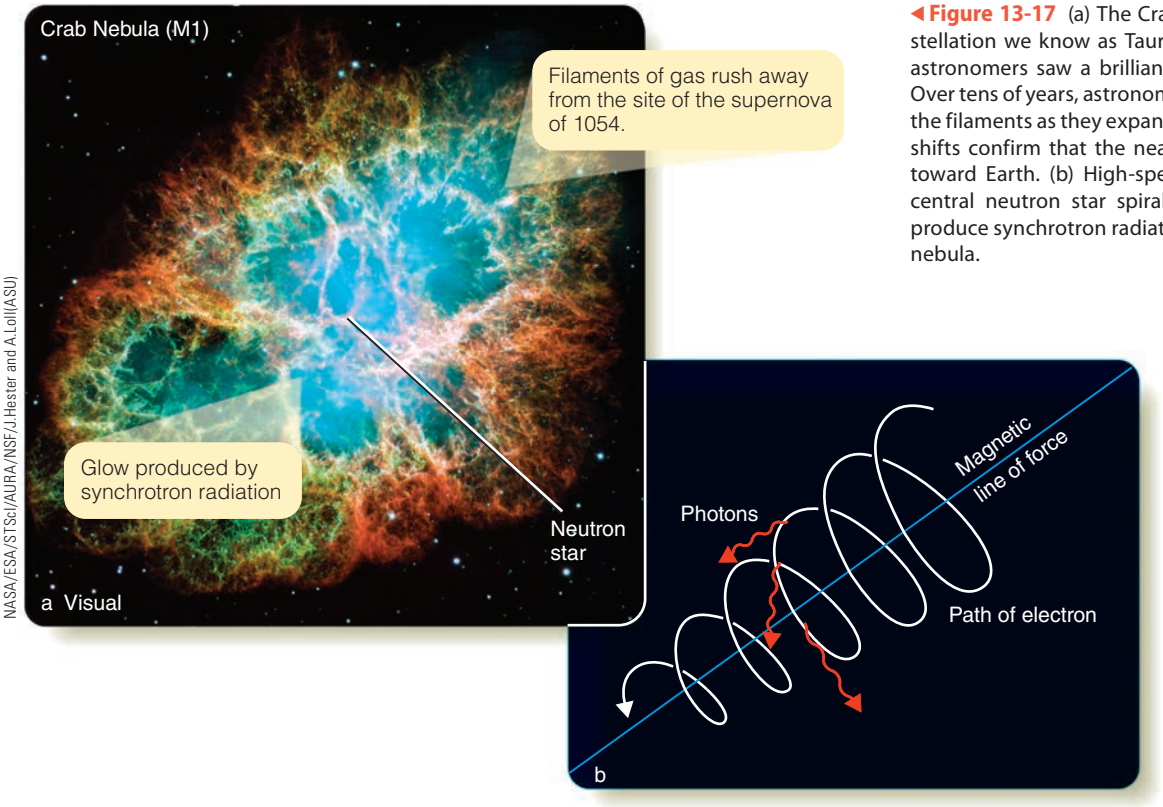
In the year 1054, Chinese astronomers saw what they called a "guest star" appear in the constellation now known as Taurus the Bull. The star quickly became so bright it was visible in the daytime. After a month's time, it slowly faded, taking almost two years to vanish from sight. Hundreds of years later, when modern astronomers turned their telescopes to the location of the guest star, they found a cloud of gas about 1.4 pc in radius expanding at 1400 km/s. Projecting the expansion back in time, they concluded that it must have begun about nine centuries ago, just when the Taurus guest star made its visit. From this and other evidence, astronomers conclude that the nebula—dubbed the Crab Nebula because in a modest-sized telescope it looks like a many-legged crab (**Figure 13-17**)—marks the site of the 1054 supernova.

The glowing filaments of the Crab Nebula appear to be excited gas flung outward by the explosion, but the hazy glow in the inner nebula is something else. Radio observations show that the gas in the nebula is emitting **synchrotron radiation**, which is electromagnetic energy radiated by high-speed electrons spiraling through a magnetic field. Electrons moving at different velocities radiate at different wavelengths, so synchrotron radiation is spread over a wide range of wavelengths. The foggy glow in the Crab Nebula is synchrotron radiation at the relatively short wavelengths of visible light, which means that the electrons must be traveling at tremendous speeds. In the nine centuries since the Crab supernova explosion, the electrons should have radiated their energy away and slowed down. There must be an energy source in the Crab Nebula that is producing very high-speed electrons. In the next chapter you will find that this is evidence there is a neutron star at the center of the Crab Nebula.

In contrast, Kepler's supernova of 1604 left nothing behind but an expanding cloud of gas and dust. Astronomers analyzing the chemical content of the cloud conclude the explosion was a type Ia supernova, so it would not have formed a neutron star or black hole. In fact, a team of astronomers reported finding the companion star that spilled mass onto the white dwarf. When the white dwarf exploded, the companion star was suddenly freed from the destroyed binary system, and its orbital velocity carried it away into space. It is seen today rushing away from the site of the explosion.

A supernova's light fades to obscurity in a year or two, but it leaves behind a remnant in the form of an expanding shell of gas, initially expelled at speeds of 10,000 km/s or more. As the supernova remnant cools, some of the gas condenses to form dust, and that makes supernovae one of the main sources of the dust in the interstellar medium. When the expanding shell of gas and dust collides with the surrounding interstellar medium, it can sweep up even more gas and excite it to produce a glowing nebula.

Supernova remnants look quite delicate and do not survive very long—a few tens of thousands of years—before they gradually mix with the interstellar medium and vanish. The Crab Nebula is a young remnant, only about 960 years old, and it isn't very large, less than 3 parsecs in diameter. Older remnants are usually larger, and some remnants are detectable only at radio and X-ray wavelengths. They have become too tenuous to emit



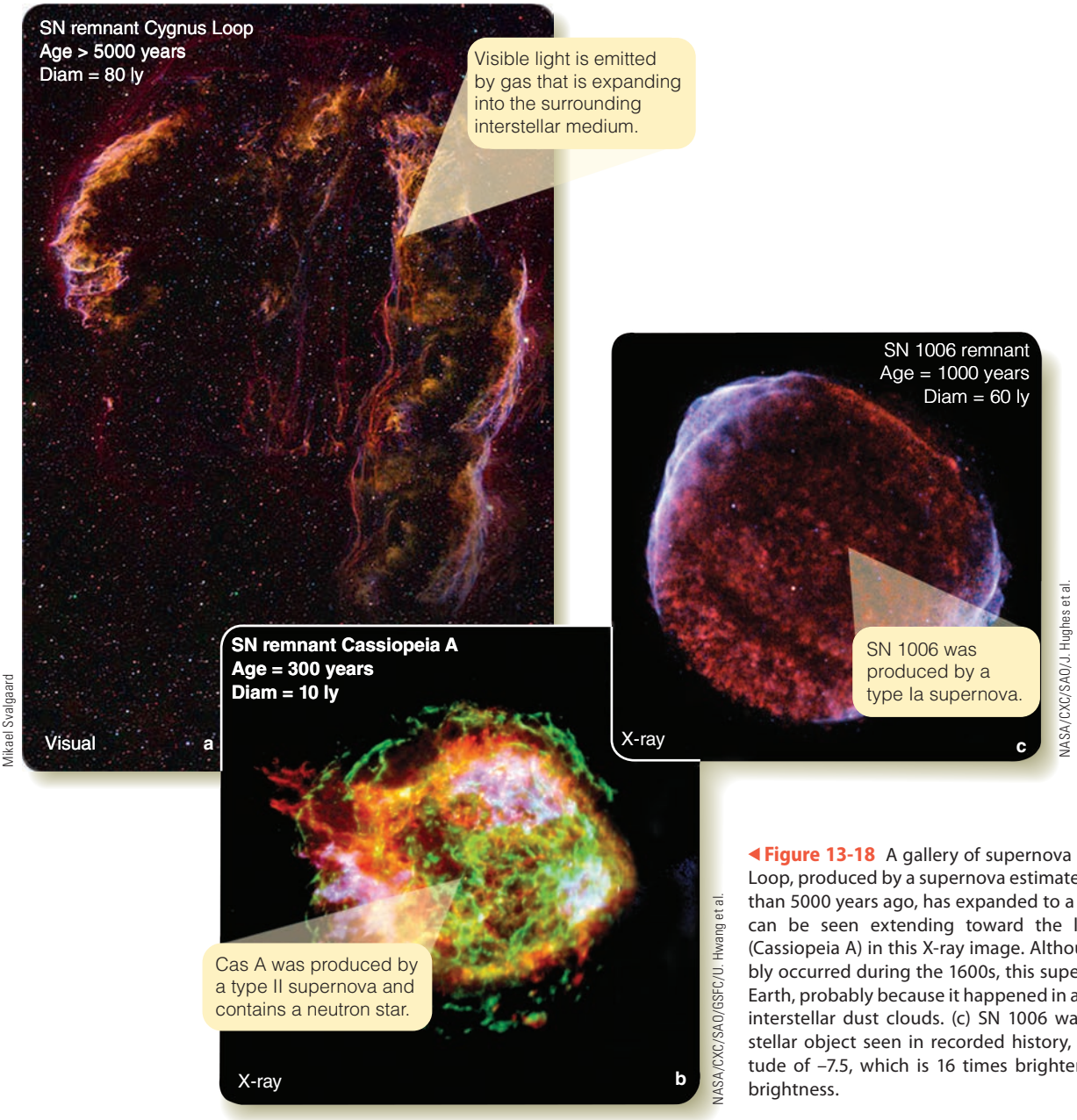
NASA/ESA/STScI/AURA/NSF/J.Hester and A.Loll(ASU)

Figure 13-17 (a) The Crab Nebula is located in the constellation we know as Taurus the Bull, just where Chinese astronomers saw a brilliant "guest star" in the year 1054. Over tens of years, astronomers can measure the motions of the filaments as they expand away from the center. Doppler shifts confirm that the near side of the nebula is moving toward Earth. (b) High-speed electrons produced by the central neutron star spiral through magnetic fields and produce synchrotron radiation, the foggy glow that fills the nebula.

much visible light, but the collision of the expanding hot gas with the interstellar medium can generate radio and X-ray radiation and allows astronomers to create images of them at these nonvisible wavelengths. In general, supernova remnants are low-density shells of gas expanding into the interstellar medium (**Figure 13-18**). (Recall from Section 11-1a that the compression of the interstellar medium by expanding supernova remnants can trigger star formation.)

The supernova remnant Cassiopeia A (Cas A) is only about 300 years old, although there is no record of a supernova being seen at that location. The *Chandra X-ray Observatory* recorded a series of X-ray images beginning in 2001 (Figure 13-18b) that allow astronomers to see the gas cloud expanding. Detailed studies reveal jets rich in silicon and iron. Infrared images made by the *Spitzer Space Telescope* combined with ground-based observations have allowed astronomers to create a three-dimensional model of the remnant. These studies show that the outer layers of the star were ejected in a spherical shell, but parts of the star's interior were ejected in an irregular, flattened disk. Theoretical models have not yet explained that geometry.

Although evidence indicates that one or two supernovae happen per year in the Milky Way Galaxy, only a few have been visible to the naked eye during all of recorded history. Arab astronomers noted one in 1006 (Figure 13-18c), and the Chinese recorded the Crab supernova in 1054. The guest stars seen in the years 185, 386, 393, and 1181 may also have been supernovae. You may recall from Sections 4-3a and d that European astronomers observed two—one in 1572 (Tycho's supernova) and one in


SN remnant Cygnus Loop
Age > 5000 years
Diam = 80 ly
Visible light is emitted by gas that is expanding into the surrounding interstellar medium.
Visual
a
Mikael Svalgaard
SN remnant Cassiopeia A
Age = 300 years
Diam = 10 ly
Cas A was produced by a type II supernova and contains a neutron star.
X-ray
b
NASA/CXC/SAO/GSFC/U. Hwang et al.
SN 1006 remnant
Age = 1000 years
Diam = 60 ly
SN 1006 was produced by a type Ia supernova.
X-ray
c
NASA/CXC/SAO/J. Hughes et al.

◀ **Figure 13-18** A gallery of supernova remnants. (a) The Cygnus Loop, produced by a supernova estimated to have exploded more than 5000 years ago, has expanded to a diameter of 80 ly. (b) Jets can be seen extending toward the left and right of Cas A (Cassiopeia A) in this X-ray image. Although the explosion probably occurred during the 1600s, this supernova was not seen from Earth, probably because it happened in a region obscured by thick interstellar dust clouds. (c) SN 1006 was probably the brightest stellar object seen in recorded history, reaching a visual magnitude of −7.5, which is 16 times brighter than Venus's maximum brightness.

1604 (Kepler's supernova). Today, most supernovae are discovered in distant galaxies, but those are faint and difficult to study.

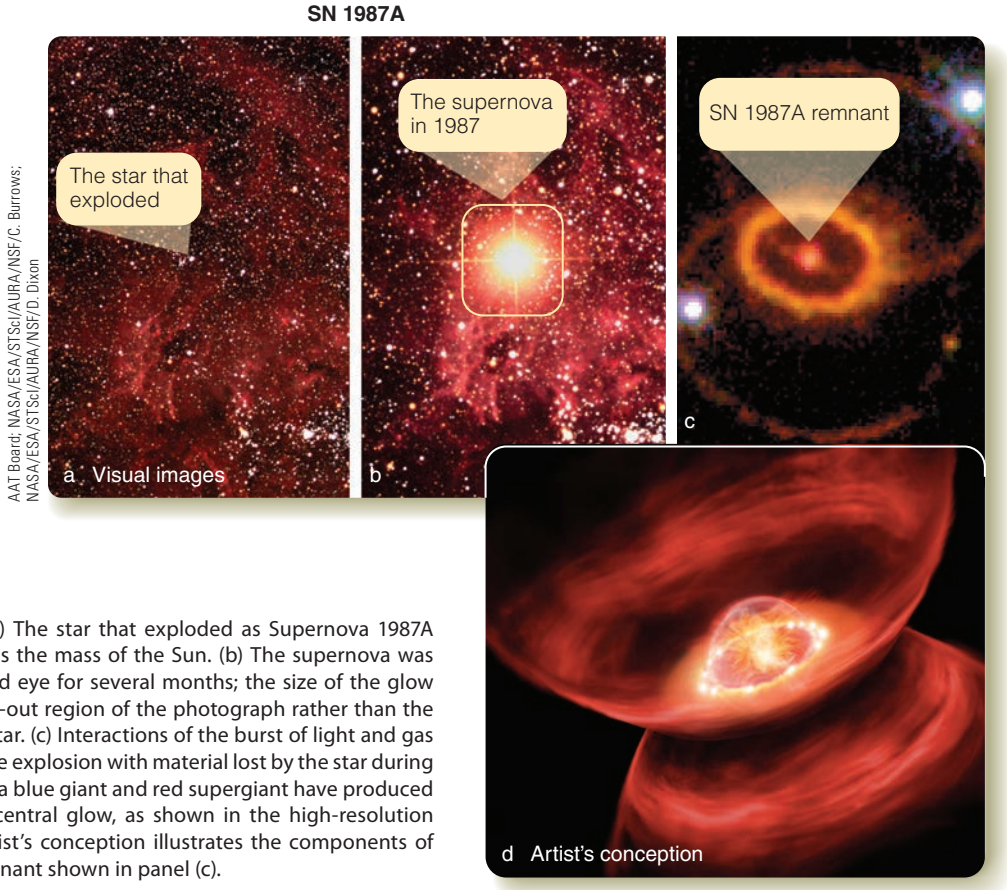
13-4f Recent Supernovae: 1987 and 2014

For the 383 years following Kepler's supernova in 1604, no naked-eye supernova was seen. Then, in 1987, the news raced around the world: Astronomers in Chile had discovered a supernova visible to the naked eye in the Large Magellanic Cloud, a small galaxy that is a companion to our Milky Way Galaxy (**Figure 13-19**). Because the supernova was only 20 degrees from the south celestial pole, it could be seen only from southern latitudes. It was named SN 1987A to denote the first supernova discovered in 1987.

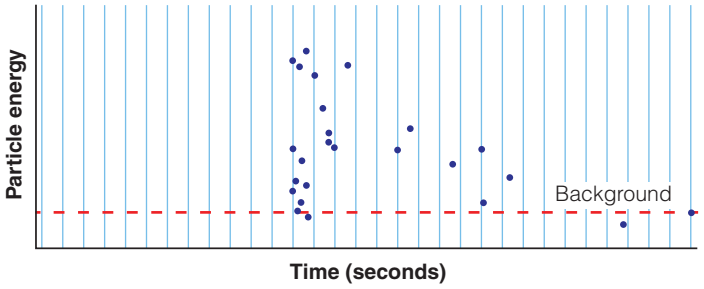
The hydrogen-rich spectrum suggested that the supernova was a type II, caused by the collapse of the core of a massive star. As the months passed, however, the light curve proved to be odd, pausing for a few weeks before rising to its final maximum (Figure 13-15). From photographs of the area made some years before, astronomers were able to determine that the star that exploded, cataloged as Sanduleak –69° 202a, was not the expected red supergiant but rather a hot, blue supergiant of only 20 $M_\odot$ and 50 $R_\odot$ (solar radii), not extreme for a supergiant. Theorists now believe that the star was chemically poor in elements heavier than helium and had consequently contracted and heated up after a phase as a cool, red supergiant, during which it had lost mass into space. The brightening of the supernova may have gone through a pause because much of the energy of the explosion went into blowing apart the smaller, denser-than-usual star.

The brightening of the expanding gases after the first few weeks seems to have been caused by the decay of radioactive nickel into cobalt, which emitted gamma-rays that heated the expanding shell of gas and made it brighter. About 0.07 $M_\odot$ of nickel was produced, equivalent to 20,000 times the mass of Earth. Gamma-rays from the decay of cobalt to iron were detected above Earth's atmosphere, and cobalt and iron spectral lines are clearly visible in the infrared spectra of the supernova.

As you learned previously, theory predicts that when a massive star's core collapses into a neutron star it should liberate a tremendous blast of neutrinos that leave the star hours before the shock wave from the interior blows the star apart. Two independent neutrino detectors—one in Ohio and one in Japan—recorded a burst of neutrinos passing through Earth at 2:35:41 AM EST on February 23, 1987, about 22 hours before the supernova was discovered (**Figure 13-20**). The data show that those neutrinos came toward Earth from the direction of the supernova. The detectors caught only 19 neutrinos during a 12-second interval, but recall that neutrinos hardly ever react with normal matter. Even though only 19 were detected, the full flood of neutrinos must have been immense. Within a few seconds of that moment, roughly 30 trillion neutrinos passed harmlessly through



▲**Figure 13-19** (a) The star that exploded as Supernova 1987A had about 20 times the mass of the Sun. (b) The supernova was visible to the naked eye for several months; the size of the glow indicates a burned-out region of the photograph rather than the actual size of the star. (c) Interactions of the burst of light and gas expanding from the explosion with material lost by the star during previous stages as a blue giant and red supergiant have produced rings around the central glow, as shown in the high-resolution image. (d) This artist's conception illustrates the components of the supernova remnant shown in panel (c).



▲**Figure 13-20** Almost a day before SN 1987A was discovered at optical wavelengths, detectors on Earth recorded 19 neutrinos arriving from the direction of the supernova. The burst dramatically exceeded the background of low-energy, sporadic neutrinos normally detected.

each human body on Earth. The detection of the neutrino blast confirms that the collapsing core gave birth to a neutron star.

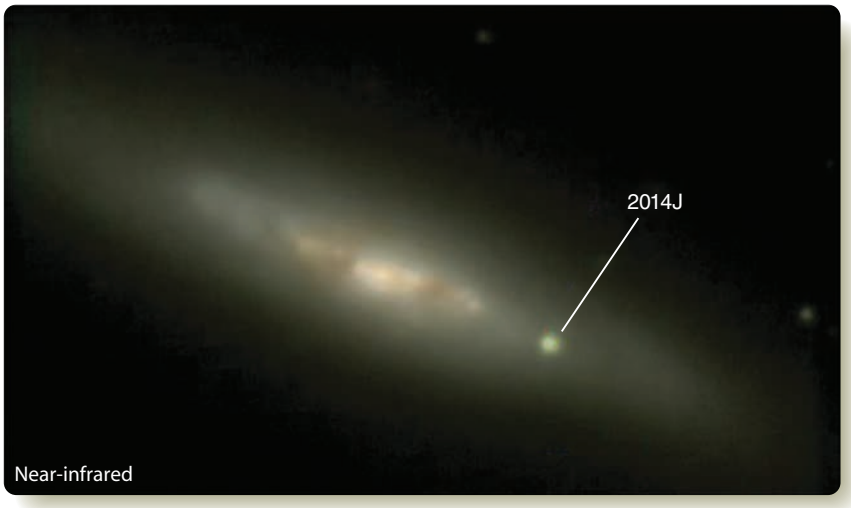
Most recently, in January 2014, a supernova was discovered in nearby galaxy Messier 82 by students in an undergraduate astronomy laboratory section at University College London. Designated SN 2014J, this supernova was determined to be of type Ia because of the lack of hydrogen lines in its spectrum. In the weeks after its discovery, a wide variety of space, airborne, and ground-based observatories observed SN 2014J in an attempt to reveal its secrets (**Figure 13-21**).

13-4g Supernovae and Life on Earth

Supernovae may seem powerful and remote, but you have a personal connection with the deaths of stars. In the previous chapter, you learned how slow-cooker nuclear reactions in medium-mass stars during their post–main-sequence life stages build atomic nuclei. This nucleosynthesis occurs in massive stars as well. When medium-mass stars die, they eject their outer layers and spread newly made carbon, nitrogen, and oxygen atoms into the interstellar medium; supernova explosions do the same. In addition, short-lived reactions in the explosion itself make other, rare elements heavier than iron such as the iodine in your thyroid gland, plus platinum, silver, gold, and uranium. Earth and we are made of atoms that were created by the stars, and the precious elements in our jewelry come exclusively from rare star explosions that happened long before Earth formed.

On the other hand, if a supernova were to occur now within a few light-years of Earth, the human race would have to abandon the surface and live below ground for at least a few decades. The burst of gamma-rays and high-energy particles from the supernova explosion could kill many life forms and cause serious genetic damage in others. Even a supernova exploding a few hundred light-years away could damage Earth's ozone layer and alter the climate. It is almost certain that, as the Sun moves through space, supernovae explode near enough to affect Earth every few hundred million years. A study of seafloor cores revealed a deposit of the isotope iron-60 in a layer that was laid down at the time of the Pliocene-Pleistocene mass extinction about 2 million years ago. Iron-60 is produced in supernova explosions and has a half-life of only 1.5 million years. To reach Earth before it decayed, this iron must have been produced in a supernova explosion no more than 100 light-years away, so some scientists have hypothesized that a nearby supernova explosion had significant effects on Earth life at that time.

Earth seems to be fairly safe at the moment. No massive red supergiant capable of exploding as a type II supernova is closer than 500 light-years away. Type Ia supernovae, however, are caused by collapsing white dwarfs, and white dwarfs are both common and hard to locate. There is no way to be sure that a white dwarf teetering on the edge of the Chandrasekhar limit does not lurk near us in space. You can take comfort, however, in the rarity of supernova explosions, which seem to occur within our part of the Galaxy only a few times per millennium. It's safe to say that Earth faces greater dangers from human ignorance than from exploding stars.



NASA/DLR/SOFIA/USRA/FLITECAM Team,UCLA/S. Shenoy

◀**Figure 13-21** An infrared image of galaxy M82 and supernova 2014J, observed from the *Stratospheric Observatory for Infrared Astronomy (SOFIA)* during the night of February 20–21, 2014. Component images at near-infrared wavelengths of 1.25, 1.65, and 2.2 microns are represented as blue, green, and red in this composite.

Practicing **SCIENCE**

Why does a type II supernova explode? To answer this type of question, a scientist focuses first on theory rather than on observations.

Models indicate that a type II supernova occurs when a massive star reaches the end of its usable fuel and develops an iron core. The iron is the final ash produced by nuclear fusion; it cannot produce energy by fusion because iron is the most tightly bound nucleus. When energy generation begins to decrease, the star contracts, but because iron can't ignite there is no new energy source to stop the contraction. Extremely rapidly, in a fraction of a second, the core of the star falls inward, and a shock wave moves outward. Aided by a flood of neutrinos and sudden convective turbulence, the shock wave blasts the star apart. Seen from Earth, the supernova brightens as its surface gases expand into space.

Now try another theoretical question: ***Why does a type Ia supernova explode?***

13-5 The End of Earth

Astronomy is actually about us. Although this chapter describes the deaths of stars, it also implicitly describes the future of our planet. There is no danger that the Sun will explode as a nova; it has no binary companion. And, as you have learned, the Sun is not massive enough to die a violent death in a supernova explosion. The Sun is a medium-mass star and will eventually die by becoming a red giant, possibly producing a planetary nebula, and finally collapsing into a white dwarf. That will mean the end of Earth.

Evolutionary models of the Sun suggest that it will survive as a star until about 7 or 8 billion years from now, but it is already growing more luminous as it fuses hydrogen into helium. In just 2 billion years, the Sun's luminosity will have increased by 20 percent, and Earth's oceans will have been evaporated to form a dense, humid atmosphere.

In about 6.5 billion years, the Sun will exhaust hydrogen in its core, begin burning hydrogen in a shell, and swell into a red giant star about 100 times its present radius. Later, helium fusion will ignite in the core, and the Sun will contract somewhat, becoming a horizontal branch ("yellow giant") star. Then, when the helium fuel is exhausted in its core and helium fusion begins in a shell, the Sun will have another episode as a red giant. The Sun's second red giant version will have a radius larger than the orbit of Earth, so that will probably mark the end of our world. Astronomers are still uncertain about some of the details, but computer models that include tidal effects predict that the expanding Sun eventually will engulf and destroy Mercury, Venus, and Earth. Even before Earth is engulfed, the Sun's increasing luminosity will certainly drive away the atmosphere and vaporize much of Earth's crust.

As a giant star, the Sun will have a strong wind and lose a substantial fraction of its mass into space. The atoms that were once in Earth's crust will be part of the expanding nebula around the Sun, and your atoms will be part of that nebula. If the white dwarf remnant Sun becomes hot enough quickly enough, it will ionize the expelled gas and light the gas (and you) up as a planetary nebula.

The most important lesson of astronomy is that we are part of the Universe and not just observers. The atoms of which Earth is made came from the interstellar medium and are destined to return to the interstellar medium in a few billion years. That's a long time, and it is possible that humans, if there are any in that far distant era, will migrate to new homes in other planetary systems. That might save humanity, but much of our planet will return to being stardust.

WHAT ARE WE?

Stardust

You are made of atoms that were cooked up inside stars. Gravity draws matter together to make stars, and although nuclear fusion delays gravity's final victory, stars must eventually die. That process of star life and star death produces atoms heavier than helium and spreads them back into the interstellar medium, where they can become part of the gas clouds that form new stars. All of the atoms in your body except for the hydrogen were made inside stars.

Some of your atoms, such as the carbon, were cooked up in the cores of medium-mass stars like the Sun and were puffed out into space when those stars died and produced planetary nebulae. Some of your atoms, such as the calcium in your bones, were made inside massive stars and were blown out into space during type II supernova explosions. Many of the iron atoms in your blood were made by the sudden fusion of carbon atoms when white dwarfs collapsed in type Ia supernova explosions. Other heavy atoms, such as iodine in your thyroid gland, selenium in your nerve cells, and gold in your class ring, were also produced in the raging violence of supernova explosions.

You are made of atoms scattered into space long ago by the violent deaths of stars. What are we? We are stardust.

Study and Review

SUMMARY

13-1 Low-Mass Stars

- Red dwarfs are stars on the main sequence of the H–R diagram that have less than about 0.5 $M_\odot$ and are completely mixed internally by convection. They cannot ignite a hydrogen fusion shell, so they cannot become giant stars. Because they have low luminosity and can fuse nearly all of their hydrogen fuel, they will remain on the main sequence for many times the present age of the Universe. When red dwarfs use up the hydrogen fuel in their cores, models indicate they should contract, increase their surface temperatures, and become bluer. Because no new energy is being generated in the core, they are expected to then cool and fade to white dwarfs. The Universe is not old enough for any white dwarfs to have formed in this way.
- Medium-mass stars with initial masses between about 0.5 and 8 $M_\odot$, including the Sun, evolve to become red giants that eventually fuse helium into carbon and oxygen. Stars with initial masses less than 8 $M_\odot$ (which, after mass loss, become red giants with less than 4 $M_\odot$) can never become hot enough to fuse carbon.
- Red giant stars lose mass rapidly into space in **superwinds**, initially by a slow wind and later a fast wind. As a giant cools, it becomes more opaque and its internal pressure increases, which causes the giant to expel the gaseous layers above the core outward in one or more surges, producing a **planetary nebula**. After the planetary nebula stage, the dying star collapses to become a white dwarf.
- White dwarfs are the remains of low- and medium-mass stars that have collapsed until they become degenerate matter. Because white dwarfs are small, dense, and generate no fusion energy, they are sometimes called **compact objects**, along with neutron stars and black holes. White dwarfs slowly cool and should eventually become **black dwarfs**. A white dwarf cannot support its own weight by its electron degeneracy pressure if its mass is greater than the **Chandrasekhar limit**, which is 1.4 $M_\odot$. Main-sequence medium-mass stars up to about 8 or 10 $M_\odot$ can lose enough mass during their lives to end as white dwarfs with masses below the Chandrasekhar limit.

13-2 The Evolution of Binary Stars

- Two stars orbiting each other control regions of space called **Roche lobes**, one around each star. The surface of the lobe is called the **Roche surface**. **Lagrange points** mark places of gravitational stability in the binary system. Only when binary stars are close enough can mass from one star flow to the other through the **inner Lagrange (L_1) point**.
- Close binary stars evolve in complex ways because they can transfer mass from one star to the other. This mass transfer explains why some binary systems contain a main-sequence star more massive than its giant companion—the Algol paradox.
- Mass transferred from one star forms an **accretion disk** around the receiving star if the receiving star is much smaller and the speed of the flow is high. Accretion disks can become hot enough to emit light, including X-rays.
- Mass transferred through an accretion disk settles onto the surface of a white dwarf and can build up a layer of fuel that, when the layer's mass exceeds a critical limit, can fuse suddenly in a **nova** explosion. A white dwarf's surface can erupt repeatedly as long as mass transfer continues to form new layers of fuel.

13-3 High-Mass Stars

- Main-sequence stars more massive than about 8 or 10 $M_\odot$ cannot lose their mass fast enough to end their lives by ejecting a planetary nebula and collapsing into a white dwarf. Such massive stars die more violent deaths.

13-4 Supernova Explosions

- Massive stars on the upper main sequence of the H–R diagram fuse nuclear fuels up to iron in their cores. These stars cannot further generate energy by nuclear fusion because iron is the most tightly bound of all atomic nuclei. When a massive star forms an iron "ash" inner core, the star no longer can support itself and it collapses, triggering a **supernova**. **Supernova remnants** are the expanding shells of hot gas ejected by supernova explosions.
- Unlike a **type II supernova**, a **type I supernova** has no hydrogen lines in its spectra. A **type Ia supernova** occurs when a white dwarf exceeds the Chandrasekhar limit, either because mass was transferred onto the white dwarf from a close binary companion or because two white dwarfs in a close binary merged to form a single white dwarf. In either process, a white dwarf's mass in the binary exceeds the Chandrasekhar limit, and the star collapses suddenly, fusing all of its carbon at once in a burst of **carbon deflagration**, which annihilates the white dwarfs. Type Ia supernovae have no hydrogen lines because white dwarfs contain little to no hydrogen.
- If an isolated aging massive star loses its hydrogen-rich atmosphere in superwinds and then collapses, the collapse could trigger a **type Ib supernova**. Another way to generate this type supernova involves a close binary: If a massive star transfers most of its hydrogen-rich outer layers to its close binary companion before collapse, then after collapse it will expel its hydrogen-poor layers in a type Ib supernova. Type Ib supernovae also lack hydrogen lines in their spectra.
- The hazy glow in the Crab Nebula is produced by **synchrotron radiation** and is evidence that some energy source remains in the nebula. Observations have found a potential source, a neutron star at the Crab Nebula's center.
- Supernova 1987A was a peculiar type II supernova explosion. Neutrinos from the explosion were detected on Earth and are evidence that the star's iron core collapsed, presumably forming a neutron star, although the neutron star has not yet been detected.
- Supernova explosions near Earth could alter Earth's ozone layer and climate. Some previous supernova explosions could have produced some of the extinctions seen in the fossil record.

13-5 The End of Earth

- The end of Earth may come in about 6.5 billion years when the Sun expands to a red giant. As the Sun evolves off the main sequence, most of the atoms now on Earth's surface are expected to blow outward and rejoin the interstellar medium. If the Sun's core collapses quickly enough into a hot white dwarf, the expelled gas will be excited to glow as a planetary nebula.

REVIEW QUESTIONS

1. Why are red dwarf star interiors thoroughly mixed?
2. Why can't the lowest-mass main-sequence stars become giants?
3. What type of remnant do the lowest-mass main-sequence stars become?
4. All white dwarfs in the Milky Way Galaxy have presumably been produced from evolved Sun-like stars and not from evolved red dwarfs. Why?
5. What evidence can you give that stars lose mass? Cite at least three observations.
6. What kind of spectrum does the gas in a planetary nebula produce? Where does the gas get the energy to radiate?
7. The coolest stars at the centers of planetary nebulae have temperatures of about 25,000 K. Why aren't planetary nebulae observed with central stars cooler than that? (*Hint:* Review Wien's law, Eq. 7-2. What band of the electromagnetic spectrum do you think is most easily able to excite emission from a nebula's gas?)
8. How are planetary nebula related to planets and nebulae, if at all?
9. Will red dwarfs pass through a planetary nebula stage? Why or why not?
10. As a white dwarf cools, it moves toward the lower right in the H–R diagram (see the right-hand page of the Concept Art: Formation of Planetary Nebulae and White Dwarfs), maintaining a constant radius. Why doesn't it contract as it cools?
11. All white dwarfs have masses in a small range—from about 0.5 to 1.4 $M_\odot$. Why?
12. Which size object best represents the size of a white dwarf—red giant, Earth, Jupiter, red dwarf, or Sun-like star?
13. Why haven't white dwarfs already cooled to black dwarfs in our Galaxy?
14. How can white dwarfs be both very hot and have very low luminosity?
15. How is a compact object different from a star?
16. What can you infer has happened if you observe a giant star in a close binary system with less mass than its main-sequence companion?
17. Why can a nova repeat whereas a supernova cannot repeat?
18. Why can't stars generate energy from iron fusion?
19. Is fusion of the lightest element in the core of a very massive star in the center, in the innermost shell, in an inner shell, in an outer shell, or in the outermost shell? What about the heaviest element?
20. What processes produce type I supernovae? Type II supernovae?
21. How do spectra show the difference between a type I supernova and a type II supernova? Why does this difference arise?
22. Consider an isolated 30-$M_\odot$ main-sequence star. How will its life end—as a type I supernova, as a planetary nebula, eroded away to nothing by a superwind, as a type II supernova, or some other fate? How do you know?
23. Consider an isolated 0.7-$M_\odot$ main-sequence star. How will its life end—as a type I supernova, as a planetary nebula, eroded away to nothing by a superwind, as a type II supernova, or some other fate? How do you know?
24. How can some supernova remnants emit X-rays?
25. **How Do We Know?** How can the ultimate causes of supernova explosions be traced back to the behavior of subatomic particles?

ACTIVE INQUIRY QUESTIONS

1. What might Earth's day and night look like from Earth's surface when the Sun becomes a white dwarf? Would Earth's day be as bright? Would we still see phases of the Moon?
2. How certain can you be that the Sun will never explode as a supernova? When scientists use the word *certainty,* what do they mean?

PROBLEMS

1. Use the formula for stellar life expectancies Eq. 12-1 to compute the life expectancy of a 0.5-$M_\odot$ star. Why might this value be an underestimate if the star is fully mixed by convection?
2. If the 0.5-$M_\odot$ red dwarf in Problem 1 fuses 100 percent of its mass from hydrogen to helium, how much energy will that star produce over its lifetime? (*Hint:* Find the energy produced per hydrogen fusion reaction in Section 8-3b.)
3. A red giant that was originally an 8-$M_\odot$ main-sequence star loses a solar mass in 100,000 years via a superwind. What is this mass loss rate in units of solar masses per year? At this mass loss rate, what will the red giant's mass be after half a million years?
4. The white dwarf star Sirius B has a visual luminosity of 0.025 $L_\odot$. If you were 1 AU from Sirius B, would that star appear brighter or fainter than the full Moon appears from Earth? (*Hint:* Information about the visual magnitudes of the Moon and the Sun, and the corresponding flux ratio, can be found in Chapter 2.)
5. Temperature on Earth is proportional to the fourth root of solar flux (T proportional to $F^{0.25}$), and flux F is proportional to the Sun's luminosity. If Sirius B, with a total luminosity of 0.055 $L_\odot$, were substituted for the Sun, what would Earth's average temperature become? (*Note:* Earth's current average temperature is 287 K.)
6. The Ring Nebula in Lyra is a planetary nebula with an angular diameter of 76 arc seconds and a distance of 5000 ly. What is its linear diameter? (*Hint:* Use the small-angle formula, Eq. 3-1.)
7. Suppose a planetary nebula is 1.0 pc in diameter, and Doppler shifts in its spectrum show that the planetary nebula is expanding at 30 km/s. How old is the planetary nebula? (*Note:* 1 pc $= 3.1 \times 10^{13}$ km and 1 yr $= 3.2 \times 10^7$ s.)
8. A planetary nebula photographed 20 years ago and then photographed again today has increased its radius by 0.6 arc second. If Doppler shifts in its spectrum show that the planetary nebula is expanding at a velocity of 20 km/s, how far away is the planetary nebula? (*Hints:* First determine how many parsecs the planetary nebula has increased in radius over 20 years. Then use the small-angle formula, Eq. 3-1.)
9. Add a fourth column to Table 13-1 and write in the atomic mass for each row's fuel element (see Appendix Table A-14). Review the curve of binding energy, Figure 8-15. Explain the trend of fusion time versus fuel atomic mass.
10. The Crab Nebula is now 1.35 pc in radius and is expanding at 1400 km/s. Approximately when did the supernova occur? (*Note:* 1 pc $= 3.1 \times 10^{13}$ km and 1 yr $= 3.2 \times 10^7$ s.)
11. The Cygnus Loop is now 2.6 degrees in diameter and lies about 500 pc distant. If it is 10,000 years old, what was its average velocity of expansion in units of km/s? (*Note:* 1 degree is 3600 arc seconds.) (*Hint:* Use the small-angle formula, Eq. 3-1, to find the linear diameter, then calculate the radius.)
12. The supernova remnant Cassiopeia A (Cas A) is expanding in radius at a rate of about 0.5 arc second per year. Doppler shifts show that the velocity of expansion is about 5700 km/s. How far away is the nebula?
13. Cassiopeia A has a radius of about 2.5 arc minutes. If it is expanding at 0.5 arc second per year, when did the supernova occur? (Note that no record can be found of a supernova being seen at that time.)
14. The distance to SN 1987A is about 170,000 ly. How long ago did the supernova occur? If the burst of neutrinos arrived at Earth 4 hours before the first light from the explosion arrived (22 hours before humans noticed it), did the neutrinos leave just before, or just after, or at the same time, as the light? (*Hint:* Reread Section 1-2, and review information about neutrinos in Section 8-3.) (*Notes*: The speed of light is 3.0×10^5 km/s; 1 ly $= 9.5 \times 10^{12}$ km.)

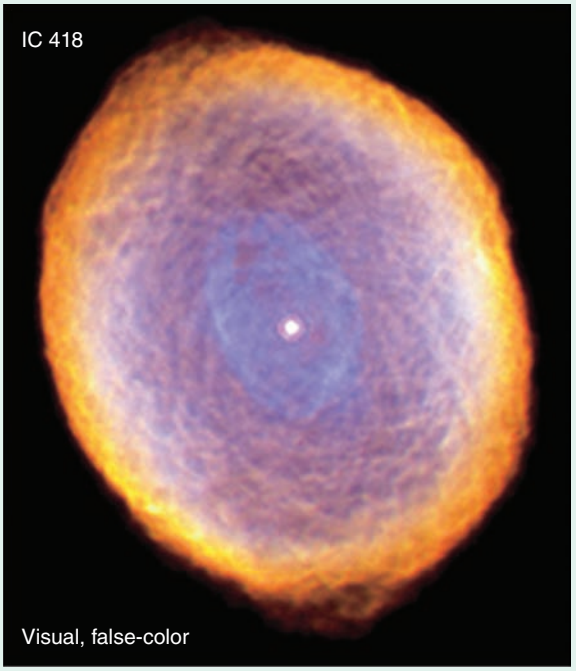
15. If 10^{58} neutrinos are expected to have carried 99 percent of the 10^{46} J of energy released in SN 1987A, how much energy of the star's collapse did each neutrino carry? If only 19 neutrinos were detected from SN 1987A, where did the other neutrinos go? If no burst of neutrinos was detected from SN 2014J, which was discovered in mid-January 2014 in the M82 galaxy, what can you conclude?

A SENSE OF PROPORTION

1. The Cat's Eye Nebula depicted in Concept Art: Formation of Planetary Nebulae and White Dwarfs is about 0.5 light-year across. The planetary region of our Solar System is about 8 light-hours across (the diameter of Neptune's orbit). If you make a model representing the Solar System using a poker chip with a diameter of about 1.5 inches, how large in that model would the Cat's Eye Nebula be?
2. Arrange the following by order of increasing energy released: type II supernova; type Ia supernova; nova eruption; the Sun's annual total energy output.

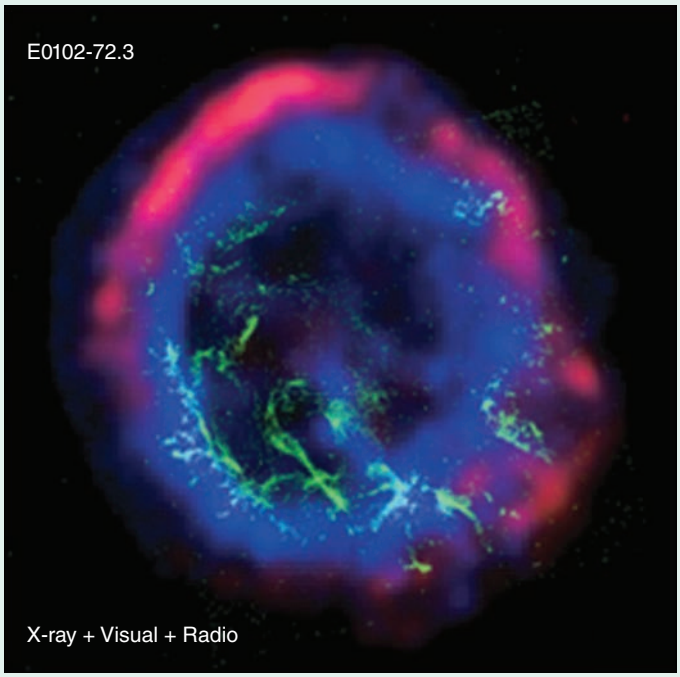
LEARNING TO LOOK

1. What processes caused a medium-mass star to produce the nebula shown here? The planetary nebula is now about 0.1 ly in diameter and still expanding. What will happen to it?



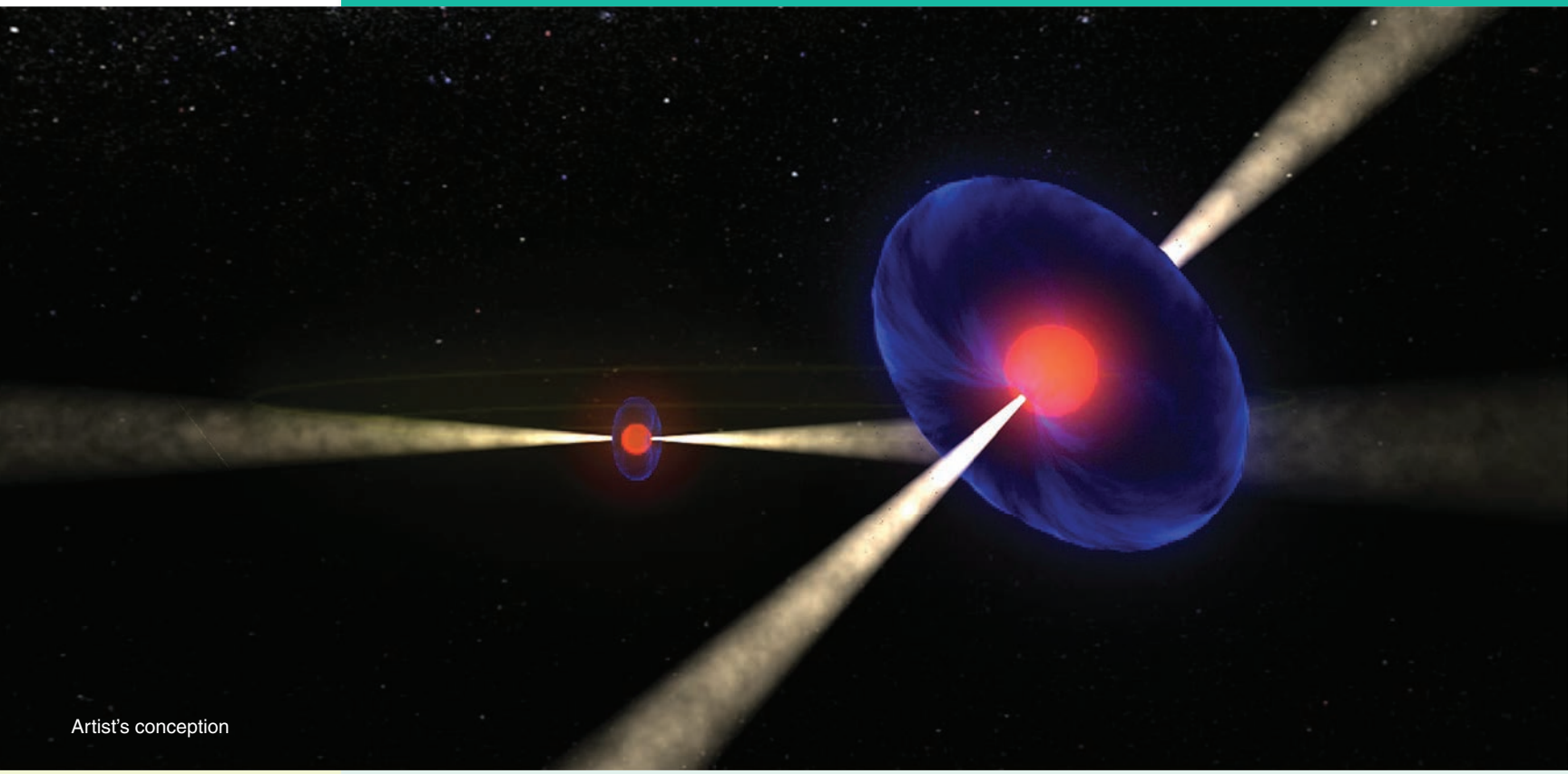
NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/ R.Sahai (JPL -Caltech) and A. Hajian (NOAO)

2. List the evolutionary stages as the Sun evolves off the main sequence to a black dwarf. For each stage, list the expected peak color (wavelength of maximum intensity) of the Sun's spectrum.
3. Look at the visual wavelength image and the combined visual plus infrared image of the Egg Nebula in the Concept Art: The Formation of Planetary Nebulae and White Dwarfs. Nebulae are typically named after the shapes they make in the visual wavelength band. What would you call this nebula if you could see it in both the infrared and the visual wavelength band? Explain which parts of the planetary nebula led you to give the planetary nebula this name.
4. Look at Figure 13-6, which is a diagram of two stars in a close binary. If an asteroid were located at the inner Lagrange point, which way would the asteroid fall: toward the star on the left side of the diagram, toward the star on the right side of the cartoon, or neither? How do you know?
5. This image combines X-ray (blue), visual (green), and radio (red) images. Observations show the sphere is expanding at a high speed and is filled with very hot gas. What do we call this object? Did a red dwarf, a medium-mass main-sequence star like the Sun, or a massive main-sequence star create this object? Roughly how old do you think this object must be?



NASA/CXC/SAO/CSIRO/ATNF/ATCA

14 Neutron Stars and Black Holes



Reprinted by permission of Daniel Cantin and McGill University office of Media Relations/ Communications

▲ Two neutron stars, collapsed remnants of massive stars, orbit each other. Both neutron stars are pulsars, meaning they emit beams of radiation that sweep around the sky as they rotate. Their orbits are edge-on as seen from Earth. The resulting eclipses allow astronomers to study the magnetic fields and gas that surround them.

In the previous two chapters, you traced the story of stars from birth to death. Now you might be asking, "What's left?" The answer depends on the mass of the star. You already know that stars with about the mass of the Sun produce white dwarfs as their remnants. More massive stars leave behind the strangest beasts in the cosmic zoo, neutron stars and black holes.

Your exploration of neutron stars and black holes will answer four important questions:

- How does theory predict the existence of neutron stars?
- What is the evidence that neutron stars really exist?
- How does theory predict the existence of black holes?
- What is the evidence that black holes really exist?

This chapter will show you more striking examples of how astronomers combine observations and theory to understand nature.

This chapter ends the story of individual stars, but it does not end the story of stars. In the next chapter, you will begin exploring the giant communities in which stars live—the galaxies.

> **Almost anything is easier to get into than out of.**
>
> AGNES ALLEN

GRAVITY ALWAYS WINS. No matter how long a star struggles to withstand its own gravity, it must eventually exhaust its fuels, collapse, and "die." Those stars that do not totally destroy themselves in that process become one of three types of compact objects—white dwarfs, neutron stars, or black holes. Almost all of the available energy has been squeezed out of compact objects, and you find them in their final, high-density states.

You studied white dwarfs in the previous chapter. Here you will learn about the most extreme of the compact objects, and you need to compare evidence and theory with great care. Theory predicts the existence of these objects, but, by their nature, they are difficult to detect. Astronomers spent decades searching for real neutron stars and black holes that could be identified as having the properties predicted by theory. It was a difficult quest but ultimately successful.

14-1 Neutron Stars

A **neutron star**, containing a little more than 1 solar mass ($M_\odot$) compressed to a radius of about 10 km, can be left as a remnant after a type II supernova explosion. To give you a sense of proportion, a white dwarf has about the same mass but is about the size of Earth, which is a bit more than 6000 km in radius. A neutron star's density is so high that physicists calculate that this material is stable only as a fluid of neutrons. Theory predicts that such an object would spin many times per second, have a surface temperature nearly as hot as the Sun's interior, and possess a magnetic field a trillion times stronger than Earth's. Two questions should occur to you immediately. First, what theory predicts such a bizarre object? And second, how can you be sure neutron stars really exist?

14-1a Theoretical Prediction of Neutron Stars

Neutron particles were discovered in a laboratory in 1932. Only two years later, Caltech astronomers Walter Baade and Fritz Zwicky published a seminal paper. They showed that some novae in historical records were much more luminous than the rest and suggested that these were caused by the collapse of a massive star's core resulting in an explosion they named a "supernova." Baade and Zwicky predicted that the core of the exploding star would form a small and tremendously dense sphere of neutrons for which they coined the term *neutron star.*

Over the following years, scientists applied the principles of quantum mechanics to see if such an object is indeed possible. Neutrons spin in ways analogous to electrons, so neutrons must also obey the Pauli exclusion principle. This means that if neutrons are packed together tightly enough, they can become degenerate just as electrons do. White dwarfs are supported by degenerate electron pressure, and quantum mechanics theory predicts that an even denser mass of neutrons could support itself by the pressure of degenerate neutrons.

How does the core of a collapsing star become a mass of neutrons? Nuclear physics provides an explanation. As the supernova explosion begins, the core collapses inward. If the collapsing core is more massive than the Chandrasekhar limit of 1.4 $M_\odot$, then it cannot reach stability as a white dwarf because the weight is too great to be supported by degenerate electrons. The collapse of the core continues, and the atomic nuclei are broken apart by gamma-rays emitted by the intensely hot material. Almost instantly, the increasing density forces the freed protons to combine with electrons and become neutrons by this reaction:

$$e + p \rightarrow n + \nu$$

A by-product of the creation of each neutron is a neutrino (symbolized by Greek lower case nu, ν). You learned in the previous chapter that the burst of neutrinos during a supernova explosion helps blast the envelope of the star away. The star's core is left behind as a neutron star.

Which stars produce neutron stars as remnants? As you learned in Section 13-1e, a star that begins its life with less than 8 to 10 $M_\odot$ can lose enough mass to end by forming a planetary nebula and leaving behind a white dwarf. More massive stars also lose mass rapidly, but model calculations indicate they cannot shed mass fast enough to reduce their mass below the Chandrasekhar limit, so it seems likely that they must die in supernova explosions. Theoretical calculations suggest that stars that begin life on the main sequence with masses in the range of 10 to 20 $M_\odot$ will leave behind neutron stars. Stars even more massive than that are thought to form black holes.

How massive can a neutron star be? That is a critical question and a difficult one to answer because scientists don't know the strength of pure neutron material. They can't make that stuff in the laboratory, so its properties must be predicted theoretically. The most widely accepted calculations suggest that a neutron star cannot be more massive than about 3 $M_\odot$. If a neutron star were more massive than that, the degenerate neutrons would not be able to support the weight; the object would collapse and presumably become a black hole. About 5 percent of neutron stars are in binary systems, allowing an estimate of their masses (look back to Section 9-5). Typical masses are about 1.4 $M_\odot$ (note that this is equal to the Chandrasekhar limit on white dwarf mass). So far the largest directly measured neutron star masses are 1.97 and 2.01 $M_\odot$, consistent with the theoretical mass limit.

Michael A. Seeds

◀**Figure 14-1** A tennis ball and a road map illustrate the relative size of a neutron star. Such an object, containing slightly more than the mass of the Sun, would fit with room to spare inside the beltway around Washington, DC.

How big are neutron stars? Mathematical models predict that a neutron star should be only about 10 km in radius (**Figure 14-1**), which, when combined with a typical mass, means it must have a density of almost 10^{15} g/cm^3. On Earth, a sugar-cube-size lump of this material would weigh more than a billion tons. This is roughly the density of an atomic nucleus, so you can think of a neutron star as matter with every bit of empty space squeezed out of it.

Simple physics—the same physics you have used in previous chapters to understand normal stars—predicts that neutron stars should be hot, spin rapidly, and have strong magnetic fields. You have seen that contraction heats the gas in a star. As gas particles fall inward, they pick up speed, and when they collide and begin to bounce around, their high infall speeds become random motions of thermal energy (look back to Section 11-1b). The sudden collapse of the core of a massive star to a radius of 10 km should heat it to a trillion (10^{12}) K. Theory predicts that a newborn neutron star should cool rapidly at first because neutrinos can escape from its entire volume and carry energy away. After a few years, the neutron star will cool down to a mere million K or so, and neutrinos won't be produced in large numbers. From that point on, the neutron star should cool slowly because energy can radiate only from the surface, and a neutron stars is so small that it has relatively little surface from which to radiate.

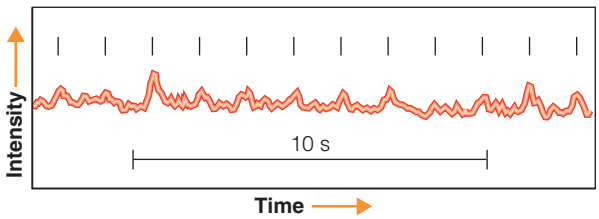
The principle of conservation of angular momentum predicts that neutron stars should spin rapidly. All stars rotate because they form from swirling clouds of interstellar matter. As a star collapses, it must rotate faster because it conserves angular momentum. Recall the example of an ice skater spinning slowly with her arms extended and then speeding up as she pulls her arms closer to her body (look back to Figure 5-6). In the same way, a collapsing star must spin faster as it pulls its matter closer to its axis of rotation. If the Sun collapsed to a radius of 10 km, its rate of rotation would increase from once every 24.5 days to more than 2000 times per second. You might therefore expect the collapsed core of a massive star to rotate at very roughly 1000 times per second.

It isn't hard to understand the prediction that a neutron star should have a powerful magnetic field. The gas of a star is ionized, and that means the magnetic field cannot move easily relative to the gas. When the star collapses, the magnetic field is caught in the gas and squeezed into a smaller volume, which can make the field as much as a billion times stronger. Because some stars start with magnetic fields more than 1000 times stronger than the Sun's, a neutron star could have a magnetic field as much as a trillion times stronger than the Sun's. That is about 10 million times stronger than any magnetic field ever produced in a laboratory.

Theory predicts the properties of neutron stars, but it also predicts that they should be difficult to observe. Neutron stars are very hot, so from your understanding of blackbody radiation (Wien's law, Eq. 7-2), you can predict they will radiate most of their energy in the gamma-ray and X-ray parts of the electromagnetic spectrum—radiation that could not be observed before the 1960s because astronomers could not yet put telescopes above Earth's atmosphere. Also, the small surface areas of neutron stars mean that they will be faint objects. Astronomers of the mid-20th century were confident that neutron stars exist even though none had yet been discovered.

14-1b The Discovery of Pulsars

In November 1967, Jocelyn Bell, a graduate student at Cambridge University in England, found a peculiar pattern in the data from a radio telescope. Unlike other radio signals from celestial bodies, this was a series of regular pulses (**Figure 14-2**). At first she



▲**Figure 14-2** The 1967 detection of regularly spaced pulses in the output of a radio telescope led to the discovery of pulsars. This record of the radio signal from the first pulsar, CP 1919, contains regularly spaced pulses (*marked by ticks*). The pulse period was later measured to be precisely 1.33730208831 seconds.

and the leader of the project, Anthony Hewish, thought the signal was interference from earthly sources, but they found it day after day in the same place in the sky. Clearly, it was celestial in origin.

Another possibility—that it came from a distant civilization—led them to consider naming it LGM, for Little Green Men. But within a few weeks, the team found three more objects in other parts of the sky pulsing with different periods. The objects were clearly natural, and the team dropped the name LGM in favor of **pulsar**—a contraction of *pulsing star.* The pulsing radio source Bell had observed with her radio telescope was the first known pulsar.

As more pulsars were found, astronomers argued over their nature. The periods ranged from 0.033 to 3.75 seconds, and each one was nearly as exact as an atomic clock. However, months of observation showed that the periods were slowly growing longer by a few billionths of a second per day. Whatever produced the regular pulses had to be highly precise, but also gradually slowing down.

It was easy to eliminate possibilities. Pulsars could not be ordinary stars or white dwarfs because neither normal stars nor white dwarfs can possibly pulse that fast. Nor could a star with a hot spot on its surface spin fast enough to produce the pulses. Even a small white dwarf would fly apart if it spun 30 times a second.

The pulses themselves gave the astronomers another clue. The pulses last only about 0.001 second, placing an upper limit on the size of the object producing the pulse. If a white dwarf with a diameter of 12,000 km blinked on and then off in that interval, you would not see a 0.001-second pulse. That's because the point on the white dwarf closest to Earth would be about 6000 km (0.02 light-seconds) closer than its center, so light from the nearest spot would arrive 0.02 s before the light from the bulk of the star. As a result, its short blink would be smeared out into a longer pulse. This is an important principle in astronomy—an object cannot change its brightness significantly in an interval shorter than the time light takes to cross its diameter. If pulses from pulsars are only 0.001 second long, then the objects cannot be larger than 300 km (200 mi) in diameter and could be even smaller. Only a neutron star is small enough to be a pulsar. In fact, a neutron star is so small that it can't blink *slowly* enough to match the observations, but it can spin as fast as 1000 times a second without flying apart.

The missing link between pulsars and rapidly spinning neutron stars was found in 1968, when astronomers discovered a pulsar at the heart of the Crab Nebula (look back to Figure 13-17). The Crab Nebula is a supernova remnant, and, as you have learned, theory predicts that some supernovae leave behind neutron stars. The short pulses and the discovery of the pulsar in the Crab Nebula were strong evidence that pulsars are neutron stars.

So far, more than 2000 pulsars have been found, but finding more can be difficult, so astronomers are using private citizens' computers to search radio data from the Arecibo Observatory in Puerto Rico for pulsing objects. You can join the search by downloading the screen saver called Einstein@Home. Then, whenever your computer is idle, it will download data files from the radio telescope and search for pulses. More than 50 new pulsars have been discovered by home computers of citizen–scientists that were coordinated by the Einstein@Home project.

14-1c A Model Pulsar

As you have now realized, scientists often work by producing a model of a natural phenomenon—not a physical model made of plastic and glue but an intellectual conception, set of equations, or computer simulation, describing how nature works in a specific instance. Models may be limited and incomplete, but they help astronomers organize their understanding.

The modern model of a pulsar has been called the **lighthouse model** and is shown in **The Lighthouse Model of Pulsars**. Notice three important points:

1. A pulsar emits beams of electromagnetic radiation that sweep around the sky as the neutron star rotates. If the beams do not sweep over Earth, the pulses will not be detectable by our telescopes.
2. The mechanism that produces the beams involves extremely high energies and is not fully understood.
3. Modern space telescopes observing from above Earth's atmosphere can make detailed images of material around young neutron stars and even locate isolated neutron stars with radiation beams that do not sweep over Earth.

Neutron stars are complicated objects with extreme conditions, and modern astronomers need to use both general relativity and quantum mechanics to try to understand them. Nevertheless, astronomers know enough to tell the life story of pulsars.

14-1d The Evolution of Pulsars

When a pulsar forms, it is spinning rapidly, perhaps hundreds of times per second. The energy it radiates into space comes from its energy of rotation, so as it blasts beams of radiation outward, its rotation slows. The average pulsar is estimated to be only a few million years old, and the oldest are about 10 million years old. Presumably, older neutron stars rotate too slowly to generate detectable beams.

You can expect that a young neutron star should emit powerful beams of radiation. The Crab Nebula is an example. Only about 960 years old, the Crab pulsar is so powerful it emits photons all across the electromagnetic spectrum, at radio, infrared, visible, X-ray, and gamma-ray wavelengths

The Lighthouse Model of Pulsars

1 Astronomers understand pulsars not as pulsing objects but rather as spinning objects emitting beams of radiation. As the pulsars spin, the beams sweep around the sky; when a beam sweeps over Earth, observers detect a pulse of radiation, like a mariner seeing a momentary flash from a spinning lighthouse beam. Understanding the details of this lighthouse model is a challenge, but the implications are clear. Although a neutron star is only a few kilometers in radius, it can produce powerful beams. Also, observers can only easily detect only those pulsars whose beams happen to sweep over Earth.

In this artist's conception, gas trapped in the neutron star's magnetic field is excited to emit light and outline the otherwise invisible magnetic field.

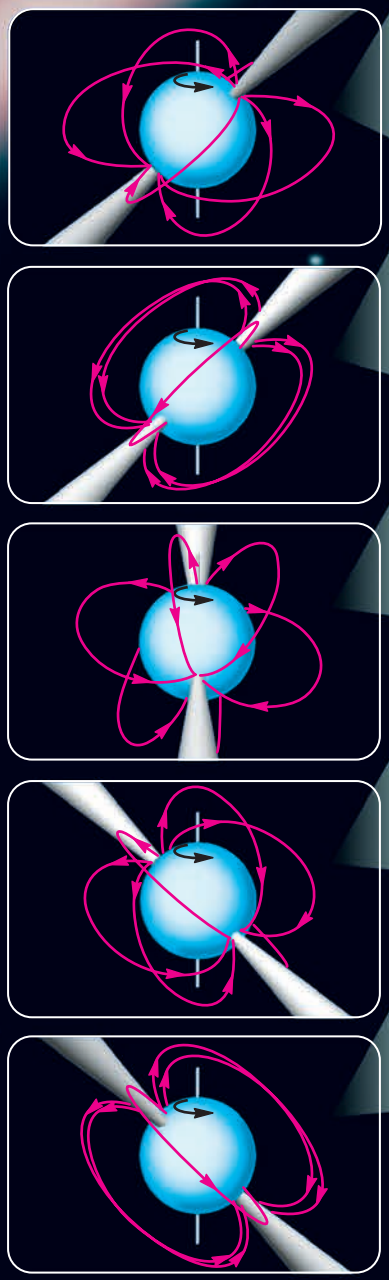
Beams of electromagnetic radiation are invisible unless they excite local gas to glow, or are pointed directly at Earth.

Artist's conception

What color should an artist use to paint a neutron star? With a temperature of a million K, the surface emits most of its electromagnetic radiation at X-ray wavelengths. Nevertheless, it would probably look blue-white to your eyes.

2 Why a neutron star emits beams of electromagnetic radiation is one of the challenging problems of modern astronomy, but astronomers have a general idea. A neutron star contains a powerful magnetic field and spins very rapidly. The spinning magnetic field generates a tremendously powerful electric field, and the field causes the production of electron–positron (matter–antimatter) pairs. As these charged particles are accelerated through the magnetic field, they emit photons in the direction of their motion, which produce powerful beams of radiation emerging from the magnetic poles of the neutron star.

Neutron Star Rotation with Beams



As is true for Earth, the magnetic axis of a neutron star can be inclined to its rotational axis.

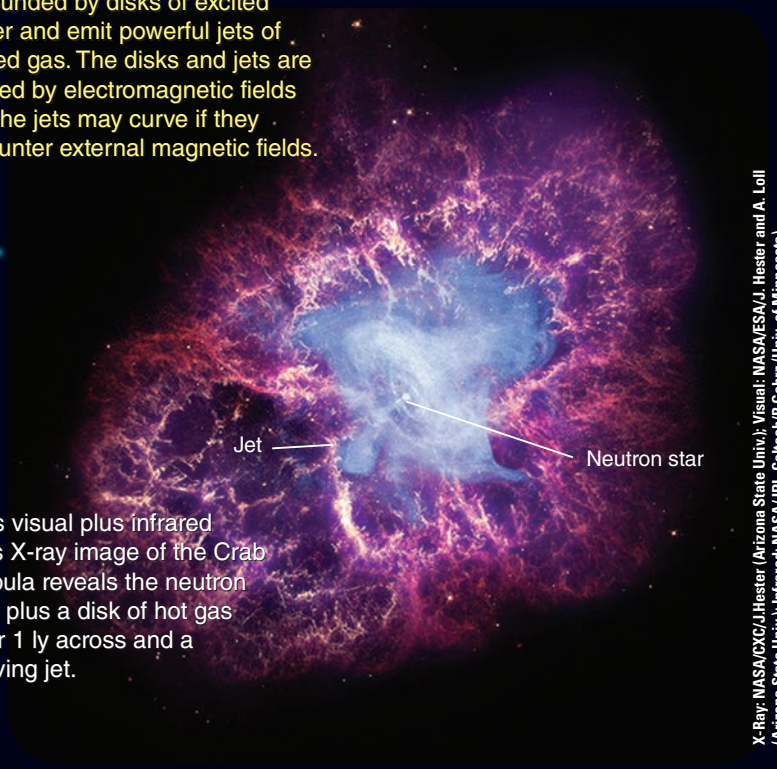
The rotation of the neutron star sweeps its beams around like lighthouse beams.

While a beam points toward Earth, observers detect a pulse.

While neither beam is pointed toward Earth, observers detect no energy.

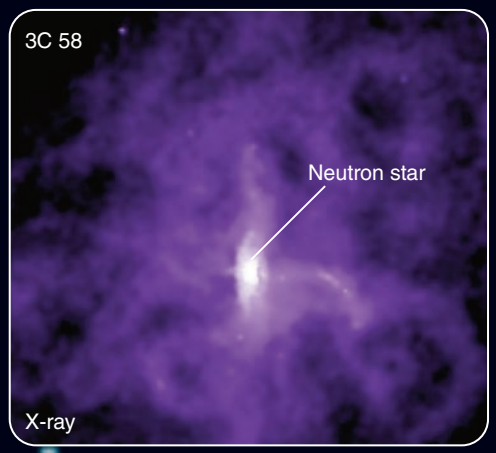
Beams may not be as exactly symmetric as in this model.

3 X-ray observations of young pulsars show that they are surrounded by disks of excited matter and emit powerful jets of excited gas. The disks and jets are shaped by electromagnetic fields and the jets may curve if they encounter external magnetic fields.



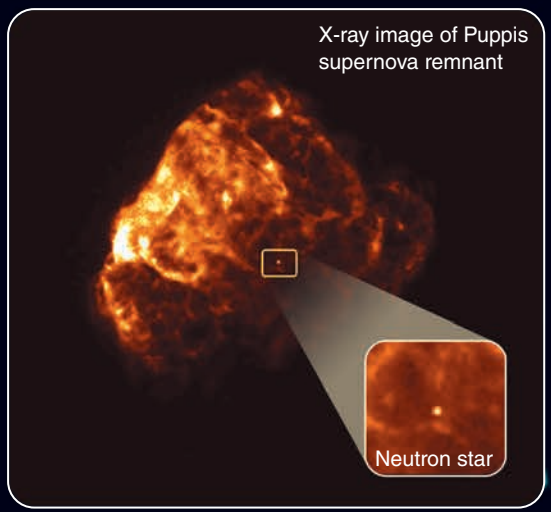
This visual plus infrared plus X-ray image of the Crab Nebula reveals the neutron star plus a disk of hot gas over 1 ly across and a curving jet.

X-Ray: NASA/CXC/J.Hester (Arizona State Univ.); Visual: NASA/ESA/J. Hester and A. Loll (Arizona State Univ.); Infrared: NASA/JPL-Caltech/R.Gehrz (Univ. of Minnesota)

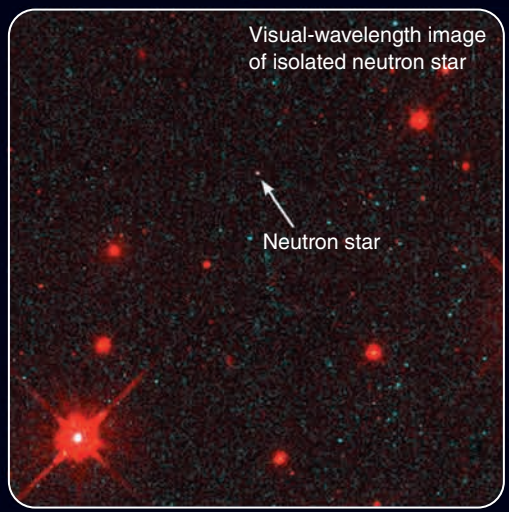


Pulsar 3C 58 at right was produced by the supernova seen on Earth in the year 1181. It pulses (rotates) 15 times per second, is surrounded by a disk, and is ejecting jets in both directions.

NASA/CXC/SAO/P. Slane et al.

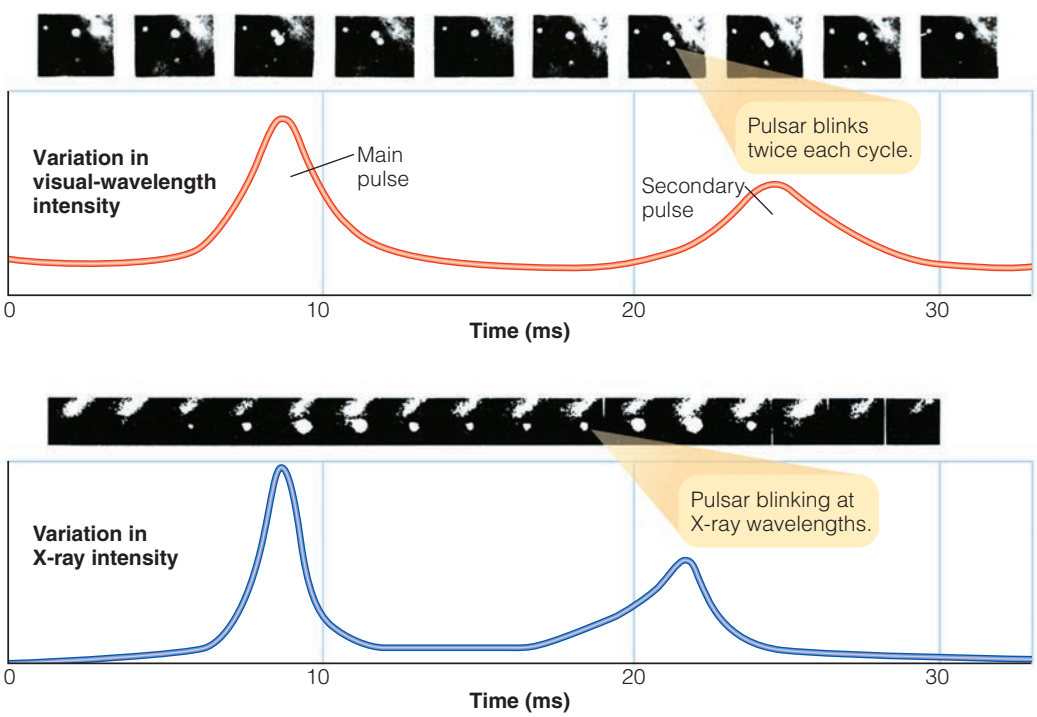


NASA/*ROSAT* Project/S. Snowden, R. Petre (GSFC), C. Becker (MIT)



NASA/ESA/STScI/AURA/NSF/F. Walter (SUNY Stony Brook)

3a If a pulsar's beams do not sweep over Earth, observers detect no pulses, and the neutron star is difficult to find. A few such objects are known, however. The Puppis A supernova remnant in the left-hand image is about 4000 years old and contains a point source of X-rays thought to be a neutron star. The isolated neutron star RX J185635-3754 seen in the right-hand image from the *Hubble Space Telescope* has a surface temperature of 700,000 K.



◀**Figure 14-3** High-speed images of the Crab Nebula pulsar show it pulsing at visual wavelengths and at X-ray wavelengths. The period of pulsation is 33 milliseconds, and each cycle includes two pulses as its two beams of unequal intensity sweep over Earth.

(**Figure 14-3**). Two orbiting gamma-ray telescopes, *Astro Rivelatore Gamma a Immagini Leggero (AGILE)* and *Fermi Gamma-Ray Space Telescope (Fermi),* have detected gamma-ray flares coming from the neutron star or from the gas nearby. This suggests rapid, violent events in that region.

Another young pulsar called the Vela pulsar (located in the Southern Hemisphere constellation Vela, the Sails) also produces visible-wavelength pulses. Compared with most pulsars, the Vela pulsar is fast, pulsing about 11 times a second. Like the Crab Nebula pulsar, it is located inside a supernova remnant. The Vela pulsar's age is estimated at a relatively young 20,000 to 30,000 years.

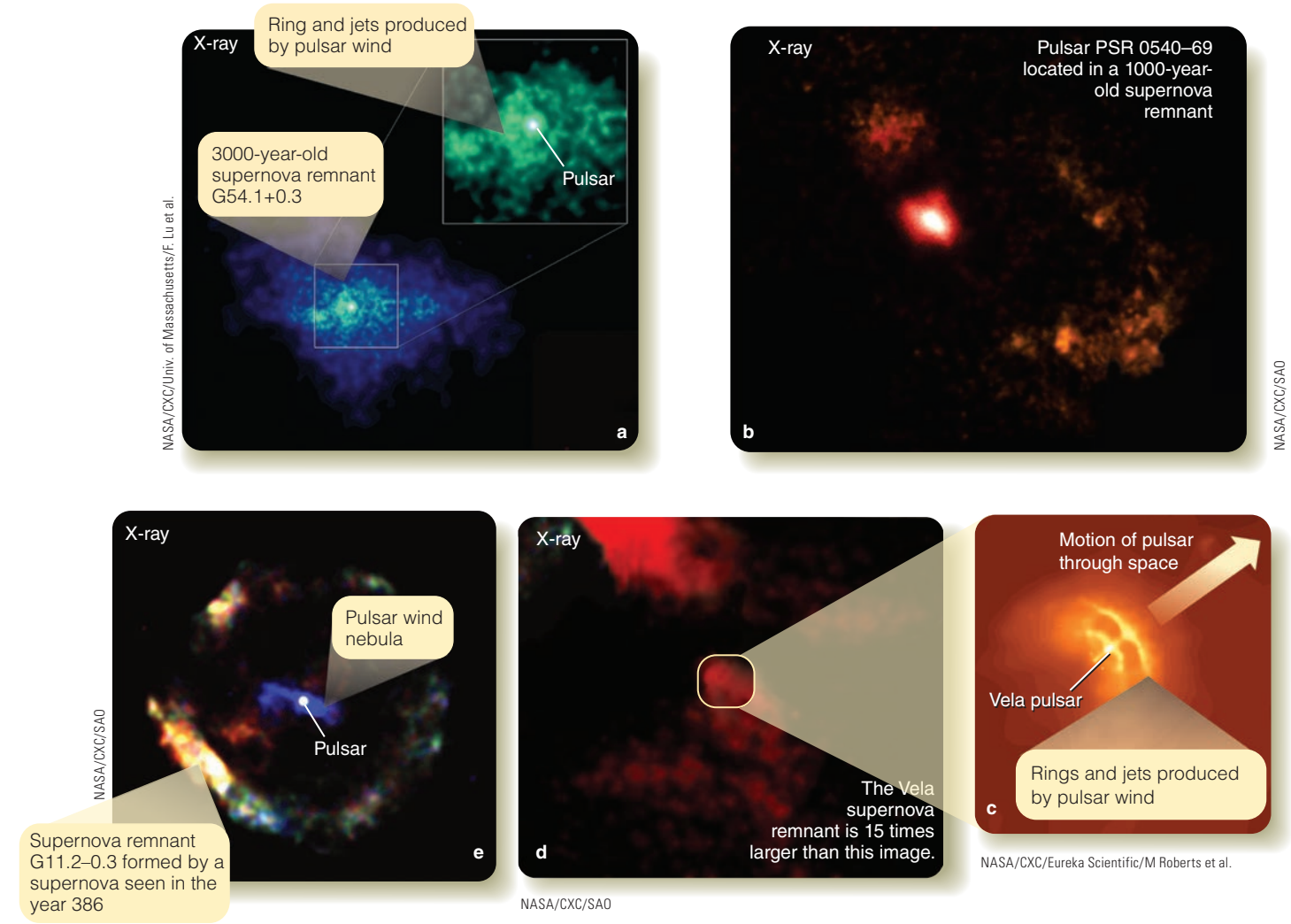
Fermi has detected examples of a new kind of pulsar that pulses only in gamma-rays. It is not clear how these pulses are produced, but because gamma-rays are extremely short-wavelength photons, the process must involve very high energies. At least one of those gamma-ray pulsars is located inside a supernova remnant that is only about 10,000 years old.

The electromagnetic energy in the beams is just a small part of the energy emitted by a pulsar. Roughly 99.9 percent of the energy flowing away from a pulsar is carried as a **pulsar wind** of high-speed atomic particles. This can produce small, high-energy nebulae near a young pulsar (**Figure 14-4**).

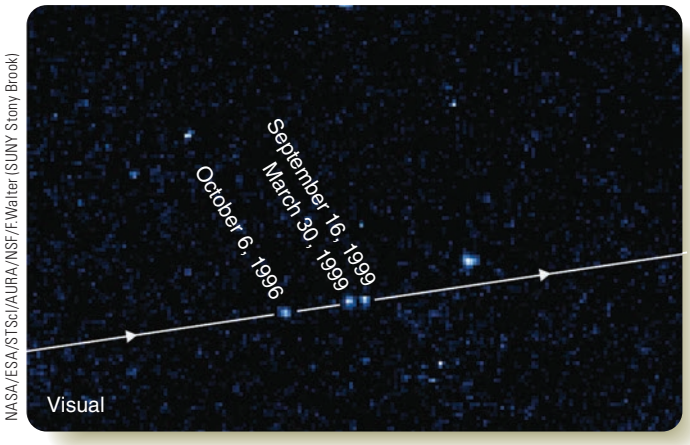
You might expect to find all pulsars inside supernova remnants, and all supernova remnants to contain pulsars, but the statistics must be examined with care. Many supernova remnants probably do contain pulsars with beams that never sweep over Earth. Also, some pulsars move through space at high velocity (**Figure 14-5**), quickly leaving their supernova remnants behind. Evidently supernova explosions can occur asymmetrically, perhaps because of the violent turbulence in the exploding core (look back to Figure 13-13), and that can kick the resulting neutron star away with a high velocity through space. Also, some supernovae probably occur in binary systems and cause the two stars to be flung apart at high velocity. (Some pulsars are known to have such high velocities that they probably can escape the disk of our Galaxy.) Also, pulsars remain detectable for 10 million years or so, but a supernova remnant cannot survive more than about 50,000 years before it is mixed into the interstellar medium. Finally, as you will learn in the next section, some supernova explosions produce black holes instead of neutron stars. For all these reasons, what is actually observed is that most pulsars are not in supernova remnants and many supernova remnants do not contain pulsars.

Astronomers conclude that the explosion of Supernova 1987A formed a neutron star because a burst of neutrinos was detected passing through Earth a few hours before the visible explosion was first detected (look back to Figure 13-20). Theory predicts that the collapse of a massive star's core into a neutron star produces a burst of neutrinos, so the detection of these neutrinos is evidence that the supernova produced a neutron star. The neutron star is predicted to be hidden at first in the center of the expanding shells of gas ejected into space, but, as the gas expands and thins, astronomers eventually should be able to detect the neutron star. Even if its beams don't sweep over Earth, astronomers expect to detect its X-ray and gamma-ray emission. Although no neutron star has yet been found in the SN 1987A remnant, astronomers continue to watch that site, waiting to see a newborn pulsar.

One reason pulsars are so fascinating is the extreme conditions found in spinning neutron stars. To see even more spectacular natural processes, you have only to look at pulsars in binary systems.



▲**Figure 14-4** The effects of pulsar winds can be seen at X-ray wavelengths. The high-energy gas of the winds is sometimes detectable, as is the interaction of the winds with surrounding gas. Not all pulsars have detectable winds.

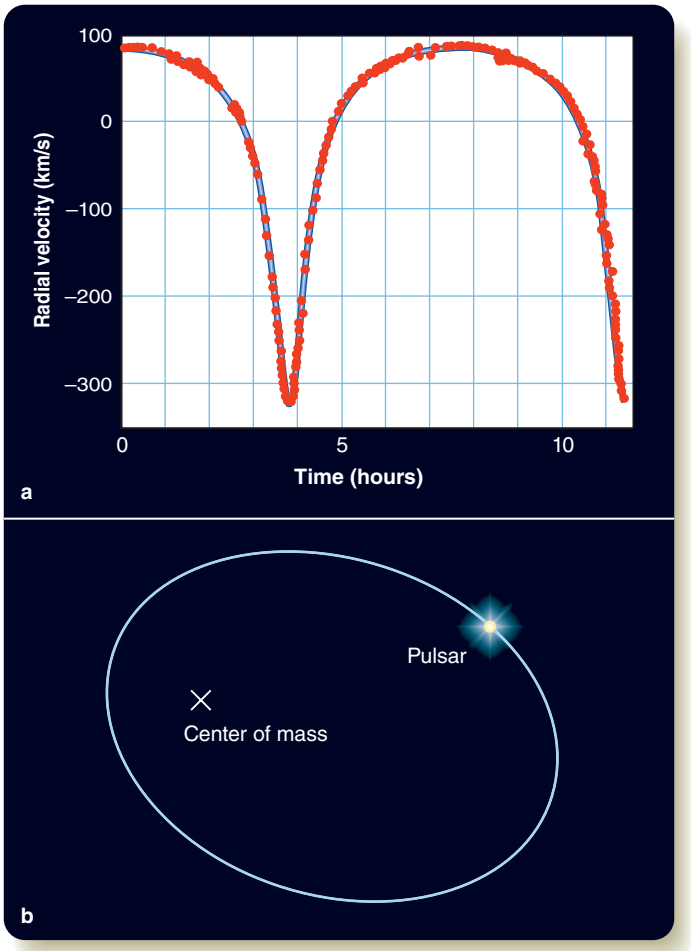


▲**Figure 14-5** Many neutron stars have high velocities through space. Here the neutron star known as RX J185635-3754 was photographed on three different dates as it rushed past background stars.

14-1e Binary Pulsars

Some pulsars are of special interest because they are located in binary systems, and astronomers can learn more about the neutron star by studying the orbital motions of the binary. Also, in some cases, mass can flow from the companion star onto the neutron star, and that produces high-energy violence.

The first binary pulsar was discovered in 1974 when astronomer Joseph Taylor and his graduate student Russell Hulse noticed that the pulse period of the pulsar PSR B1913+16 was changing. The period first grew longer and then grew shorter in a cycle that took 7.75 hours. Thinking of the Doppler shifts seen in spectroscopic binaries, the radio astronomers realized that the pulsar had to be in a binary system with an orbital period of 7.75 hours. When the orbital motion of the pulsar carries it away from Earth, astronomers see the pulse period lengthen slightly—a redshift. Then, when the pulsar rounds its orbit and approaches



▲**Figure 14-6** (a) The radial velocity of pulsar PSR B1913+16 can be found from the Doppler shifts in its pulsation. (b) Analysis of the radial velocity curve allows astronomers to determine the pulsar's orbit. Here the center of mass does not appear to be at a focus of the elliptical orbit because this orbit is inclined to the line of sight from Earth.

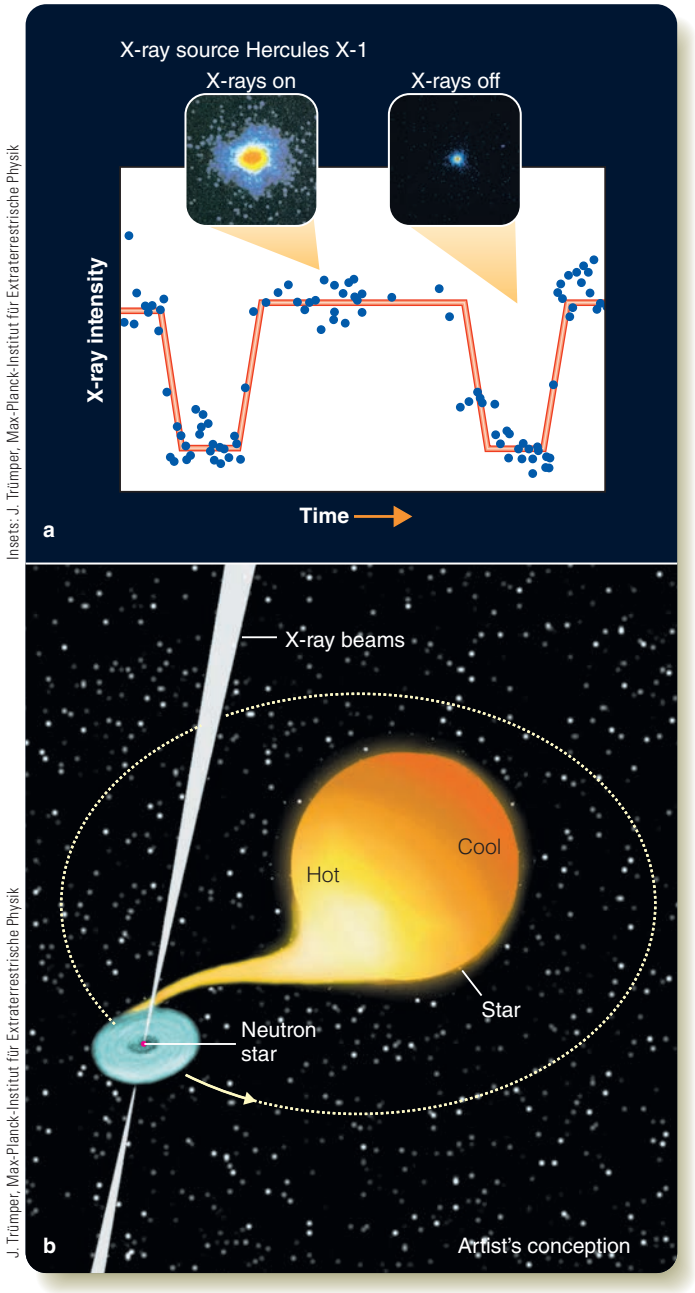
Earth, they see the pulse period shorten slightly—a blueshift. From these changing Doppler shifts, Taylor and Hulse calculated the radial velocity of the pulsar as it moves around its orbit just as if it were a spectroscopic binary star. The resulting graph of radial velocity versus time was then analyzed to find the shape of the pulsar's orbit (**Figure 14-6**). These studies of PSR B1913+16 showed that the binary system consists of two neutron stars, one of which is "silent" (not pulsing as seen from Earth), separated by a distance roughly equal to the radius of our Sun.

Yet another surprise was hidden in the motion of PSR B1913+16. Einstein's general theory of relativity describes gravity as a curvature of space-time. Einstein realized that any rapid change in a gravitational field should spread outward at the speed of light as **gravitational radiation**. Taylor and Hulse were able to show that the orbital period of the binary pulsar is slowly growing shorter because the stars are radiating orbital energy away as gravitational radiation and gradually spiraling toward each other. (Normal binary stars are too far apart and orbit too slowly to emit significant gravitational radiation.) Taylor and Hulse won the Nobel Prize in 1993 for their work using binary pulsars to confirm general relativity. In 2016, it was announced that the Laser Interferometer Gravitational Wave Observatory, LIGO, had made the first direct detection of gravitational radiation.

In 2004, radio astronomers announced the discovery of a double pulsar: two pulsars that orbit each other in only 2.5 hours. The spinning beams from both pulsars sweep over Earth (look back to the image that opens this chapter). One spins 44 times a second, and the other spins once every 2.8 seconds. This system is a pulsar jackpot because the orbits are nearly edge-on to Earth, and the powerful magnetic fields and the gas trapped in the fields eclipse each other, giving astronomers an opportunity to study their size and structure. Furthermore, the theory of general relativity predicts that these pulsars are emitting gravitational radiation so that their separation is decreasing by 7 mm per day. The two neutron stars will merge in 85 million years, presumably triggering a violent explosion. In the meantime, the steady decrease in orbital period can be measured, giving astronomers another test of general relativity and gravitational radiation.

In addition to producing gravitational radiation, a neutron star's intense gravitational field means that binary pulsars can be sites of tremendous violence if matter is transferred from a star to a neutron star. The gravitational field is so strong that an astronaut stepping onto the surface of a neutron star would be instantly smooshed into a layer of matter only one atom thick. Matter falling into this gravitational field releases titanic amounts of energy. If you dropped an apple onto the surface of a neutron star from a distance of 1 AU, it would hit with an impact equivalent to a half-megaton nuclear warhead. In general, a particle falling from a large distance to the surface of a neutron star will release energy equivalent to about 0.2 mc^2, where m is the particle's mass at rest and c is the speed of light. Even a small amount of matter flowing from a companion star to a neutron star can generate high temperatures and release X-rays and gamma-rays.

Hercules X-1 is an example of such an active system, containing neutron star and a 2-$M_\odot$ main-sequence star that orbit each other with a period of 1.7 days (**Figure 14-7**). (Its name means it was the first X-ray source found in the constellation Hercules.) Matter flowing from the normal star into an accretion disk around the neutron star reaches temperatures of millions of kelvins and emits a powerful X-ray glow. Interactions of the gas with the neutron star's magnetic field produce beams of X-rays that sweep around with the rotating neutron star (Figure 14-7b). Earth receives a pulse of X-rays every time a beam points this way. The X-rays shut off completely every 1.7 days when the neutron star is eclipsed behind the normal star. Hercules X-1 has many different high-energy processes going on simultaneously, but this quick sketch serves to illustrate how complex and powerful such binary systems can be during mass transfer.



▲**Figure 14-7** (a) Sometimes the X-ray pulses from Hercules X-1 are on, and sometimes they are off. A graph of X-ray intensity versus time looks like the light curve of an eclipsing binary. (b) In Hercules X-1, matter flows from a star into an accretion disk around a neutron star producing X-rays, which heat the near side of the star to 20,000 K compared with only 7000 K on the far side. X-rays turn off from Earth's point of view when the neutron star is eclipsed behind the star.

The X-ray source 4U 1820-30 illustrates another way neutron stars can interact with other stars. In this system, a neutron star and a white dwarf orbit their center of mass with a period of only 11 minutes (**Figure 14-8**). The separation between the two objects is only about one-third the distance between Earth and the Moon—smaller than a main-sequence star. To explain how such a close pairing of stellar remnants could originate from what once was an ordinary binary star system, theorists suggest that a neutron star collided with a giant star and went into an orbit inside the star. (Recall the low density of the outer envelope of giant stars.) The neutron star would gradually eat away the giant star's envelope from the inside, and the star's core eventually collapsed into a white dwarf. Matter still flows from the white dwarf into an accretion disk and then down to the surface of the neutron star (Figure 14-8c), where it accumulates in a degenerate layer until it ignites helium fusion to produce a burst of X-rays. Objects called **X-ray bursters** are thought to be such binary systems involving mass transferred to a neutron star, and the bursts repeat each time a large enough layer of degenerate fuel accumulates. Notice the similarity between this mechanism and that responsible for nova explosions (look back to Figure 13-8).

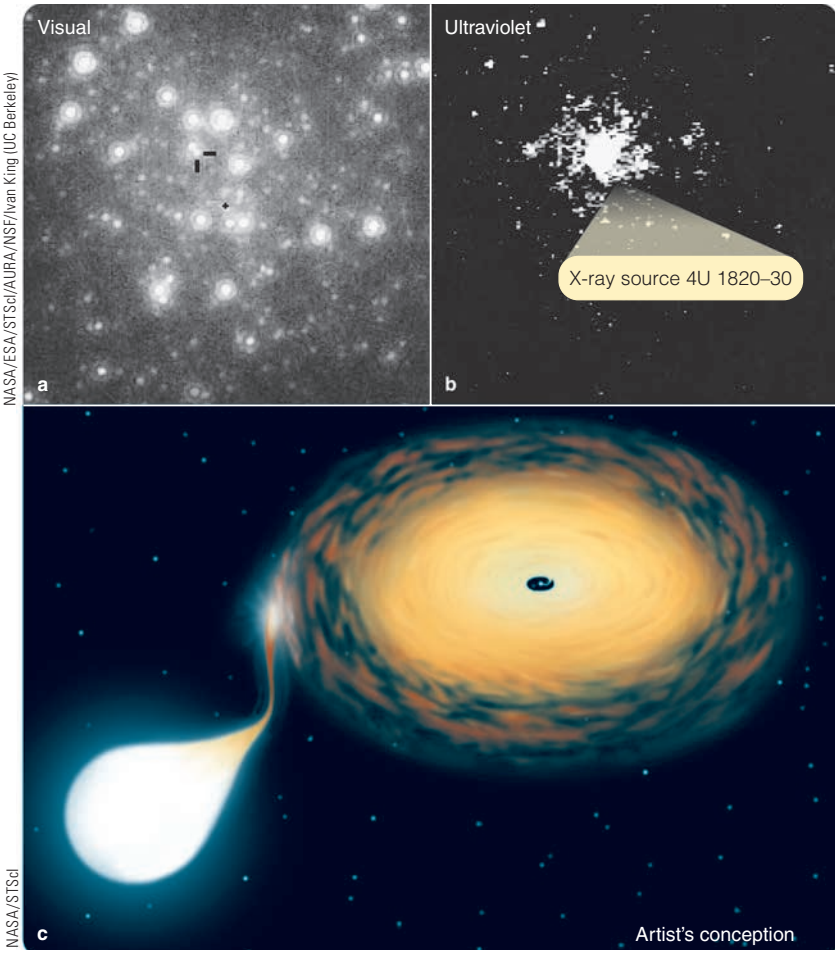
14-1f The Fastest Pulsars

Your knowledge of pulsars suggests that newborn pulsars should blink rapidly, and old pulsars should blink slowly. In fact, the handful that blink the fastest may be quite old. The currently fastest known pulsar is cataloged as PSR J1748−2446ad. It pulses 716 times a second and is slowing down only slightly. The energy stored in the rotation of a neutron star at this rate is equal to the total energy of a supernova explosion, so it seemed difficult at first to explain this pulsar. It now appears that the speedy pulsar is an old neutron star that has gained mass and rotational energy from a companion in a binary system. Like water hitting a mill wheel, the matter falling on the neutron star has spun it up to that fantastic rate. With its weak magnetic field, it slows down only gradually and will continue to spin rapidly for a very long time.

A number of other very fast pulsars have been found. They are generally known as **millisecond pulsars** because their pulse periods, and therefore their periods of rotation, are around a millisecond (0.001 s). This rapid rotation produces some fascinating physics. If a neutron star 10 km in radius spins 716 times a second, then its equator must be traveling at almost a quarter the speed of light. That's fast enough to flatten the neutron star into an oblate shape.

A significant number of these speedy pulsars have been detected by gamma-ray telescopes in orbit above Earth's atmosphere. It may be that all millisecond pulsars produce gamma-rays because their extremely rapid rotation causes high-energy interactions between pulsar magnetic fields and nearby gas.

The hypothesis that millisecond pulsars were spun up by mass transfer from a companion star is plausible, but scientists demand evidence, and evidence has been found. One example, the X-ray source XTE J1751-305, is a pulsar with a period of 2.3 milliseconds. X-ray observations show that it is gaining mass from a companion star. The orbital period is only 42 minutes, and the mass of the companion star is 0.014 $M_{\odot}$, only 15 times the mass of Jupiter. The evidence suggests that this neutron star has devoured all but the last morsel of its binary partner.



◀**Figure 14-8** (a) At visible wavelengths, the center of star cluster NGC 6624 is crowded with stars. (b) In the ultraviolet, one object stands out, an X-ray source consisting of a neutron star orbiting a white dwarf. (c) An artist's conception shows matter flowing from the white dwarf into an accretion disk around the neutron star.

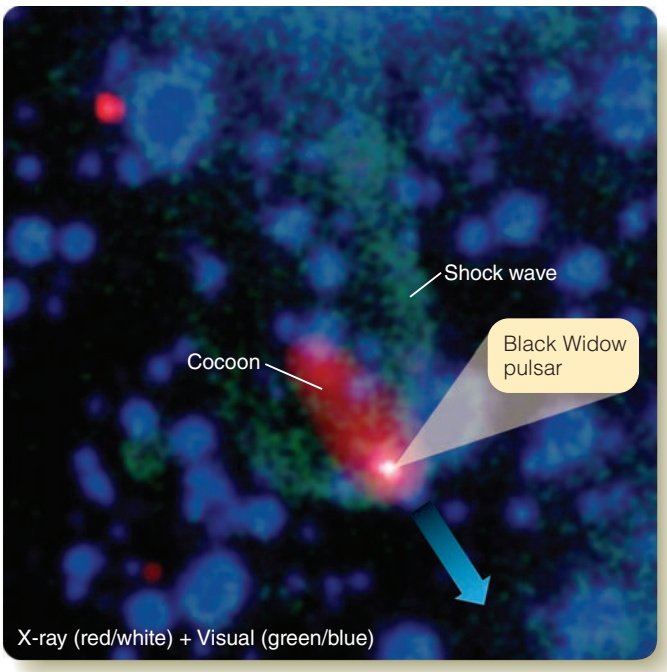
14-1g Pulsar Planets

Because a pulsar's period is so precise, astronomers can detect tiny variations by comparing their observations with atomic clocks. When astronomers checked pulsar PSR B1257+12, they found variations in the period of pulsation much like those caused by the orbital motion of a binary pulsar (**Figure 14-10a**). However, in the case of PSR B1257+12, the variations were much smaller, and when they were interpreted as Doppler shifts, it became evident that the pulsar was being orbited by at least two objects with planetlike masses of 4.3 and 3.9 times the mass of Earth—only about 1/100,000 the mass of the Sun. The gravitational tugs of these planets make the pulsar wobble about the center of mass of the system by less than 800 km, producing the observed tiny changes in period (**Figure 14-10b**).

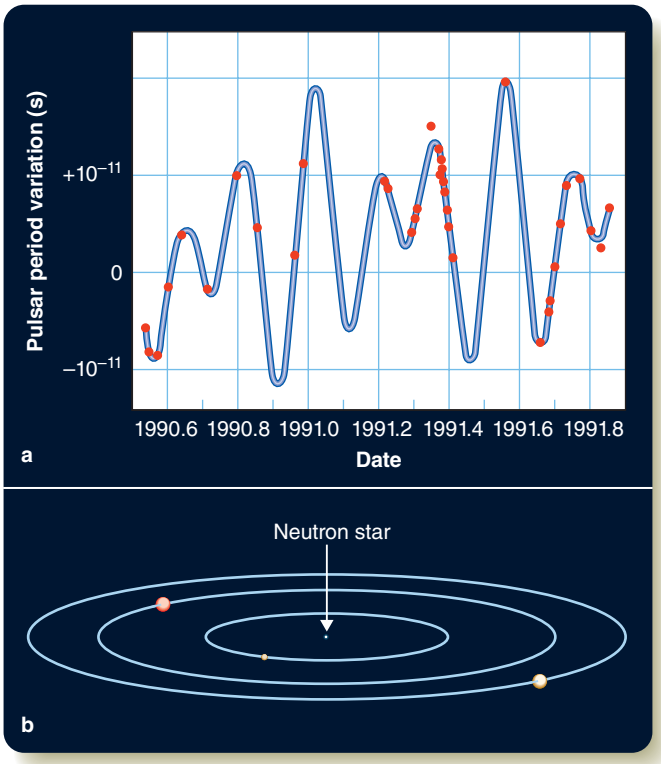
Astronomers greeted this discovery with both enthusiasm and skepticism. As usual, they looked for ways to test the hypothesis. Simple gravitational theory predicts that planets in the same system should interact and slightly modify each other's orbits. When the data were analyzed, that interaction was found,

Astronomers wondered about a few millisecond pulsars that do not have companions. Were they produced by some process other than mass transfer? Pulsar PSR B1957+20, also known as the Black Widow (**Figure 14-9**), seems to show that they all might once have been members of binary systems. The Black Widow has a period of 1.6 milliseconds and is orbited by a brown dwarf companion. There is no evidence of current mass transfer. However, spectra of the system show that blasts of radiation and high-energy particles from the neutron star are now evaporating the companion. When the companion is completely gone, presumably a new solitary millisecond pulsar will be left behind. Since the Black Widow pulsar's discovery, 20 more examples of pulsars vaporizing low-mass mates have been found.

"Show me," say scientists, and in the case of neutron stars, astronomers had gathered enough evidence to be confident that such objects really do exist. Hypotheses are less firm about the formation and evolution of neutron stars, and about how they emit beams of radiation, but continuing observations at many wavelengths are expanding astronomers' understanding of these last remnants of massive stars. In fact, precise observations have turned up some objects no one expected.


Shock wave
Black Widow pulsar
Cocoon
X-ray (red/white) + Visual (green/blue)
X-ray: NASA/CXC/SAO/ASTRON/B. Stappers et al.; Visual: AAO/J. Bland-Hawthorn & H. Jones

▲**Figure 14-9** The Black Widow pulsar and its companion star are moving rapidly through space, creating a shock wave like the bow wave in front of a speedboat. The shock wave confines high-energy particles shed by the pulsar into an elongated cocoon (*red*).



▲ **Figure 14-10** (a) The dots in this graph represent observations showing that the period of pulsar PSR B1257+12 varies from its average value by a fraction of a billionth of a second during the time span of about one Earth year. The blue line indicates the variation predicted by a model of the effects of planets orbiting the pulsar. (b) As the planets orbit the pulsar, they cause it to wobble by less than 800 km (500 mi), a distance that is invisibly small in this diagram. The outermost planet orbits about 0.46 AU from the pulsar, farther than Mercury but closer than Venus is from the Sun.

further confirming the existence of the planets. In fact, later data revealed the presence of a third planet with only 1/40 the mass of Earth, about twice the mass of Earth's Moon. This illustrates the astonishing precision of studies based on pulsar timing.

You might wonder how a neutron star can have planets. You're in good company; astronomers find this puzzling as well. The three innermost planets that orbit PSR B1257+12 are closer to the pulsar than Venus is to the Sun. Any planets that orbited a star that closely would have been absorbed or vaporized when the star expanded to become a supergiant. Furthermore, if the supernova explosion that produced the neutron star didn't destroy planets outright, it would have allowed the planets to escape from their orbits by suddenly reducing the mass of the star. So, how can these planets exist? One possibility is that these planets are the remains of a stellar companion that was devoured by the neutron star. In fact, PSR B1257+12 spins very fast (161 pulses per second), suggesting that it was spun up in a binary system. However, the *Spitzer Space Telescope* observing at infrared wavelengths has detected a ring of gas and dust around a different rapidly spinning neutron star. If supernova explosions somehow can leave such rings of material behind, then perhaps planets can form from the accumulation of matter in the rings.

PSR B1257+12 is not unique. Another planet has been found orbiting a pulsar that is part of a binary system with a white dwarf in a very old star cluster. The characteristics of this system indicate, however, that the planet may have been captured rather than being debris from the supernova explosion that made the neutron star. Planets probably orbit other neutron stars, and small shifts in the timing of the pulses may eventually reveal their presence.

What might these worlds be like? Formed from the remains of evolved stars, they might have chemical compositions much richer in heavy elements than Earth. You can imagine visiting these worlds, landing on their surfaces, and hiking across their valleys and mountains made of gold, lead, and uranium. Above you, the neutron star would glitter in the sky, a tiny but deadly point of light.

Practicing SCIENCE

Why, if you want to detect neutron stars, might you want to use an X-ray telescope? A scientist needs to be able to decide which instrument is needed to answer a given question.

First, recall that a neutron star is very hot because of the heat released when it contracts from the size of a supergiant stellar core down to a radius of 10 km. It could easily have a surface temperature of 1 million K, and Wien's law (Eq. 7-2) tells you that such an object will radiate most intensely at a very short wavelength—X-rays or gamma-rays. Normal stars are much cooler and emit only weak X-rays unless they have hot accretion disks. At visual wavelengths, stars are bright, and neutron stars are faint, but at X-ray wavelengths, the neutron stars stand out from the crowd.

Now pretend you are seeking funds for a research project. ***What observations would you propose to determine whether a newly discovered pulsar is young or old, single or a member of a binary system, and alone or accompanied by planets?***

14-2 Black Holes

You have studied white dwarfs and neutron stars, two of the three end states of dying stars. Now it's time to consider the third end state—black holes.

Although the characteristics of black holes are difficult to discuss without using sophisticated mathematics, simple physics indicates that they might exist, and our knowledge of stellar evolution along with a mass limit for neutron stars implies that they should exist. The problem is to confirm that they are real. What objects observed in the Universe could actually be black holes? More difficult than the search for neutron stars, the quest for black holes has nevertheless met with success.

To start your study of black holes, consider a simple question. How fast must an object travel to escape from the surface of a celestial body?

14-2a Escape Velocity

Suppose you threw a baseball straight up. How fast must you throw it if it is not to come down? Of course, gravity will always

pull back on the ball, slowing it, but if the ball is traveling fast enough to start with, it will never come to a stop and fall back. Such a ball will escape from Earth.

In Section 5-2c and the Concept Art: Orbits, you learned that the escape velocity is the initial velocity an object needs to escape from a celestial body. Whether you are discussing a baseball leaving Earth or a particle escaping a collapsing star, escape velocity depends on two things: The mass of the celestial body and the distance from the center of mass to the escaping object. If the celestial body has a large mass, its gravity is strong and you need a high velocity to escape, but if you begin your journey farther from the center of mass, the velocity needed is less.

Using Equation 5-1b, you calculated that the escape velocity to leave Earth's surface is 11.2 km/s (25,100 mph), but if you could launch spaceships from the top of a tower 1000 miles high, the escape velocity would be only 10.4 km/s (23,300 mph). Using the same equation, you can calculate the escape velocity from the surface of the Sun. The Sun's mass is 1.99×10^{30} kg and its radius is 6.96×10^{8} m:

$$V_e = \sqrt{\frac{2GM_\odot}{R_\odot}} = \sqrt{\frac{2 \times 6.67 \times 10^{-11} \times 1.99 \times 10^{30}}{6.96 \times 10^{8}}}$$

$$= \sqrt{3.81 \times 10^{11}} = 6.18 \times 10^{5}\ \text{m/s} = 618\ \text{km/s}$$

The Reverend John Mitchell, a British astronomer, realized one especially interesting consequence of Newton's laws of gravity and motion. If an object is massive enough or small enough, its escape velocity can be greater than the speed of light. In 1783, Mitchell pointed out that an object 500 times the radius of the Sun but of the same density would have an escape velocity greater than the speed of light. Then, "all light emitted from such a body would be made to return towards it." Such a dense object could never be seen because light cannot leave it. Mitchell didn't know it, but he was talking about a black hole. Now we know from Einstein's special theory of relativity that nothing can travel faster than the speed of light, so if light cannot escape an object, nothing else can, either.

14-2b Schwarzschild Black Holes

If the core of a star contains more than 3 $M_\odot$ when it collapses, no force can stop it. It can't stop collapsing when it reaches the density of a white dwarf because degenerate electrons cannot support that weight, and it can't stop when it reaches the density of a neutron star because not even degenerate neutrons can support that weight. The object must collapse to zero radius. No force remains to stop the object from collapsing to zero radius.

As an object collapses, its density and the strength of its surface gravity increase. If an object collapses to zero radius, its density and gravity become infinite. Mathematicians call such a point a **singularity**, but in physical terms it is difficult to imagine an object of zero radius. Some theorists say that a singularity is impossible in the real Universe and that the laws of quantum physics will somehow halt the collapse at a subatomic size estimated to be approximately 10^{20} times smaller than a proton. Astronomically, it seems to make little difference.

If the contracting core of a star becomes small enough, the escape velocity in the region of space around it is so large that no light can escape. This means you can receive no information about the object or about the region of space near it. Because it emits no light, such a region is called a **black hole**. If the core of an exploding star collapsed to become a black hole, the expanding outer layers of the star could produce a supernova remnant, but the core would vanish without a trace.

To further understand black holes, you need to consider relativity. As you learned in Section 5-3b, Einstein published a mathematical theory of space and time in 1916 that became known as the general theory of relativity. Einstein treated space and time as a single entity called space-time, and his equations showed that gravity could be described as a curvature of space-time. Almost immediately, the astronomer Karl Schwarzschild found a way to solve Einstein's equations to describe the gravitational field around a single, nonrotating, electrically neutral lump of matter. That solution included the first general relativistic description of a black hole, and nonrotating, electrically neutral black holes are now known as Schwarzschild black holes. In recent decades, theorists such as Roy Kerr and Stephen Hawking have found ways to apply the sophisticated mathematical equations of the general theory of relativity and quantum mechanics to describe rotating and electrically charged black holes. The differences are minor at the level of this discussion, so you can consider all black holes as Schwarzschild black holes.

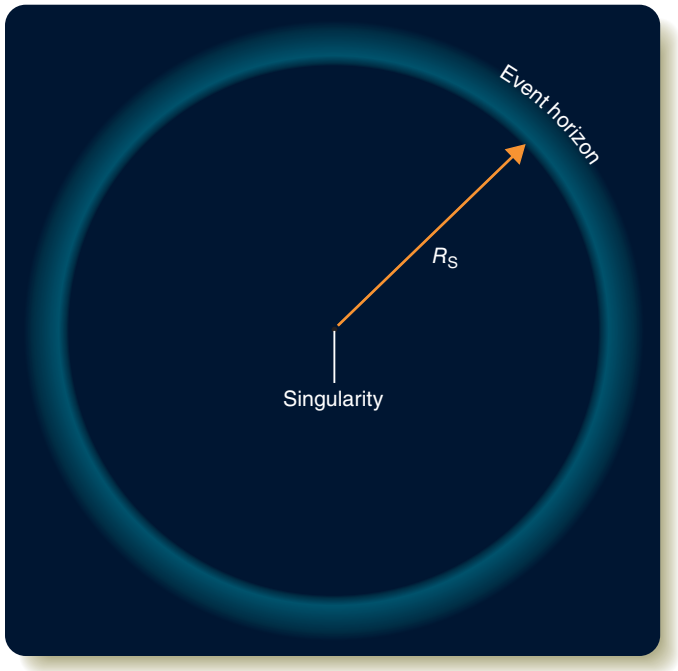
Schwarzschild's solution shows that if matter is packed into a small enough volume, then space-time curves back on itself. Objects can follow paths that lead into the black hole, but no path leads out, so nothing can escape. Because not even light can escape, the inside of the black hole is totally beyond the view of an outside observer. The **event horizon** is the boundary between that isolated volume of space-time and the rest of the Universe. The radius of the event horizon is called the **Schwarzschild radius, R_S**. A collapsing stellar core that is massive enough must unavoidably shrink inside its Schwarzschild radius to produce a black hole (Figure 14-11).

Although Schwarzschild's work was highly mathematical, his conclusion is quite simple. The Schwarzschild radius depends only on the mass of the object:

KEY EQUATION 14-1 The Schwarzschild Radius

$$R_s = \frac{2GM}{c^2}$$

In this simple formula, G is the gravitational constant, M is the mass (in kilograms), and c is the speed of light (in meters per



▲ **Figure 14-11** A black hole forms when an object collapses to a small size (perhaps to a singularity) and the escape velocity becomes so great light cannot escape. The boundary of the black hole is called the event horizon because any event that occurs inside is invisible to outside observers. The radius of the black hole R_S is the Schwarzschild radius.

second). ***Example:*** Calculate the Schwarzschild radius for the mass of the Sun, 1.99×10^{30} kg. ***Solution:***

$$R_s = \frac{2GM_\odot}{c^2} = \frac{2 \times 6.67 \times 10^{-11} \times 1.99 \times 10^{30}}{(3.00 \times 10^8)^2}$$

$$= 2.95 \times 10^3 \text{ m} = 2.95 \text{ km}$$

Thus, a 1-$M_\odot$ black hole has a Schwarzschild radius of 3 km, a 10-$M_\odot$ black hole has a Schwarzschild radius of 30 km, and so on, linearly proportional to the mass (**Table 14-1**).

Every mass has a corresponding Schwarzschild radius. For example, an object with the mass of Earth has a Schwarzschild radius of about 1 cm, meaning that Earth would become a black hole if you could squeeze it smaller than that radius. Of course, Earth will not collapse spontaneously to become a black hole because the strength of the rock and metal in its interior is more than sufficient to support its weight. Only extinguished stellar cores more massive than about 3 $M_\odot$ can form black holes under the sole influence of their own gravity.

Table 14-1 The Schwarzschild Radius

	Mass, $M_\odot$	R_S
Star	10	30 km
Star	3	9 km
Sun	1	3 km
Earth	0.000003	0.9 cm

It is a **Common Misconception** to think of black holes as giant vacuum cleaners that will eventually suck in everything in the Universe. A black hole is just a gravitational field, and at large distances, its gravity is the same as that of a normal object with the same mass. If the Sun were to be suddenly replaced by a 1-$M_\odot$ black hole, the orbits of the planets would not change at all. **Figure 14-12** illustrates this by representing gravitational fields as curvature of the fabric of space-time. Normal uncurved space-time is represented by a flat plane, and the presence of a mass such as a star curves the plane to produce a depression. The extreme curvature around a black hole produces a deep funnel-shaped surface in this graphic representation. You can see from the graphs that the gravity of a black hole becomes extreme only when you approach it closely.

Now you can check off another **Common Misconception** that may strike you as silly. Because of special effects in movies and TV, some people think black holes should actually look like funnels. Of course, the graphs of the strength of gravity around black holes look like funnels, but black holes themselves are not shaped like funnels. If you could approach a black hole, you might be able to see hot gas swirling inward, but you wouldn't be able to see the black hole itself.

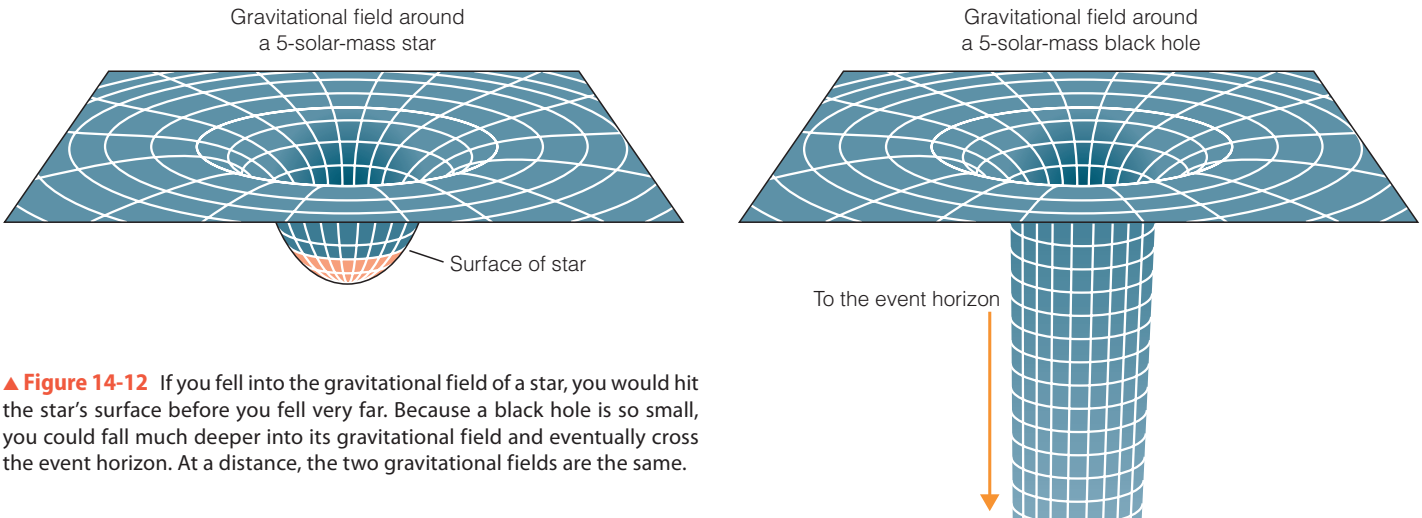
14-2c Leaping into a Black Hole

Before you can search for real black holes, you need to understand theoretical predictions about what happens near a black hole. To explore those ideas, you can imagine leaping, feet first, into a Schwarzschild black hole.

If you were to leap into a black hole with the mass of the Sun from a distance of 1 AU, the gravitational pull at your starting point would not be very large, and initially you would fall slowly. Of course, the longer you fell and the closer you came to the center, the faster you would travel. Your wristwatch would tell you that you fell for about two months by the time you reached the event horizon.

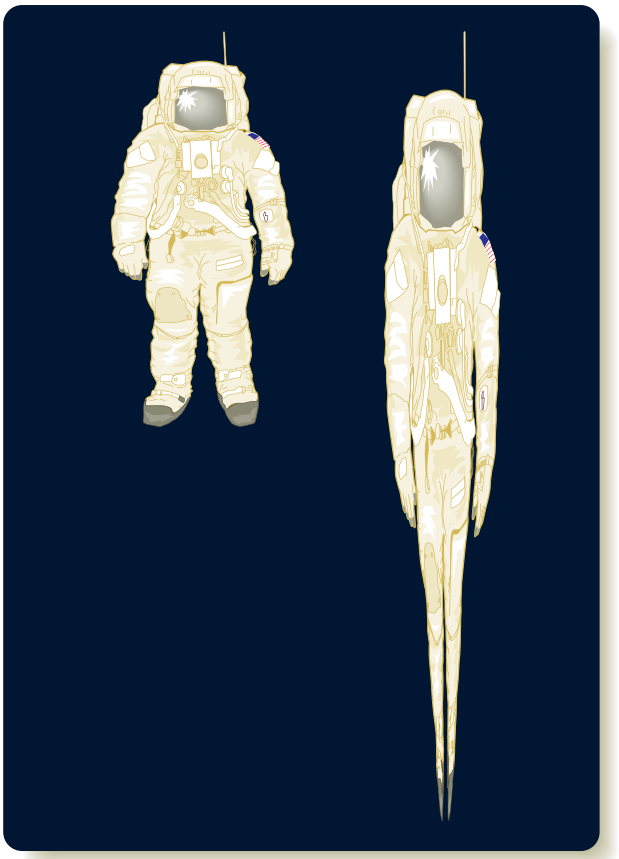
Your friends who stayed behind would observe something different. They would see you falling more and more slowly as you came closer to the event horizon because, as predicted by general relativity and confirmed by experiments, time slows down in curved space-time. This is known as **time dilation**. In fact, your friends would never actually see you cross the event horizon. To them you would fall more and more slowly until you seemed hardly to move. Generations later, your descendants could focus their telescopes on you and see you still inching closer to the event horizon. You, however, would have sensed no slowdown and would conclude that you had reached the event horizon after two months.

Another relativistic effect would make it difficult to see you with normal telescopes. As light travels out of a gravitational field, it loses energy, and its wavelength grows longer. This is known as the **gravitational redshift**. Your friends would need to observe at longer and longer wavelengths to detect you.



▲ **Figure 14-12** If you fell into the gravitational field of a star, you would hit the star's surface before you fell very far. Because a black hole is so small, you could fall much deeper into its gravitational field and eventually cross the event horizon. At a distance, the two gravitational fields are the same.

Although these relativistic effects seem merely peculiar, other effects would be quite unpleasant. If you were falling feet first, you would feel your feet, which would be closer to the black hole, being pulled in more strongly than your head. This is a tidal force, and at first it would be minor. But as you got closer to the black hole, the tidal force would become very large. Another tidal force would compress you as both your left and your right side were pulled on paths converging toward the center of the black hole. For any black hole with a mass like that of a star, the tidal forces would crush you sideways and stretch you lengthwise before you reached the event horizon (**Figure 14-13**).



The friction from such severe and rapid distortions of your body would heat you to millions of kelvins, and you would emit X-rays and gamma-rays as you approached the event horizon. (Needless to say, these effects would render you inoperative as a careful observer.)

Some people have suggested that it is possible to travel through the Universe by jumping into a black hole in one place and popping out of another somewhere far across space. That might make for good science fiction, but tidal forces would make it an unpopular form of transportation even if it worked. You would certainly be separated from your luggage.

Imagining a leap into a black hole is not entirely frivolous. Now you know how to find a black hole: Look for a strong source of X-rays. It might be a black hole into which matter is falling and being heated to high temperatures.

14-2d The Search for Black Holes

Do black holes really exist? The first X-ray telescopes reached space in the 1960s, allowing orbit in the 1970s, allowing astronomers to begin searching for evidence of black holes. They tried to find one or more objects that were obviously black holes. That very difficult search is a good illustration of how the unwritten rules of science help scientists understand nature (**How Do We Know? 14-1**).

A black hole alone is totally invisible because nothing can escape from its event horizon, but if matter flows into a black hole, it will whirl through an accretion disk and become hot enough to emit X-rays before it disappears down the hole. An

◀ **Figure 14-13** Leaping feet first into a black hole. A person of normal proportions (*left*) would be distorted by tidal forces (*right*) long before reaching the event horizon around a typical black hole of stellar mass. Tidal forces would stretch the body lengthwise while compressing it laterally. Friction from this distortion would heat the body to high temperatures.

Checks on Fraud in Science

How do you know scientists aren't just making stuff up?

The unwritten rules of science make fraud difficult, and the way scientists publish their research makes it almost impossible. Scientists depend on each other to be honest, but they also double-check everything.

For example, all across North America, black-capped chickadees sing the same quick song. Some people say it sounds like *Chick-a-dee-dee-dee,* but others say it sounds like *Hey-sweetie-sweetie-sweetie.* You could invent tables of data and publish a paper reporting that you had recorded chickadees around Ash Lake in northern Minnesota and that they sing a backward song: *Sweetie-sweetie-sweetie-hey.* Experts on brain development and animal learning would be amazed, and your research might secure you praise from your colleagues, a job offer at a prestigious university, or a generous grant—but only if you could get away with it.

The first step in your scheme would be to publish your results in a scientific journal. Because the journal's reputation rests on the accuracy of the papers it publishes, the editor sends all submitted papers to one or more experts for peer review. Those world experts on chickadees would almost certainly notice things wrong with your made-up data tables. On their recommendation, the editor would probably refuse to publish your paper.

Even if your fake data fooled the peer reviewers, you would probably be found out once the paper was published. Experts on bird song would read your paper and flock to Ash Lake to study the strange backwards bird songs themselves—only to find out that your report was not correct. By the next spring, you would be found out—and the journal would be forced to publish an embarrassing retraction of your article.

One of the rules of science is that good results must be repeatable. Scientists routinely repeat the work of others, not only to check the results but also as a way to start a new research topic. When someone announces a new discovery, other scientists begin asking, "How does this fit with other observations? Has this been checked? Has this been peer reviewed?" Until a result has been published in a peer-reviewed journal, scientists treat it with extra care. In fact, NASA and other federal science agencies won't schedule a news conference to report results from one of their scientists, no matter how exciting, until a peer-reviewed journal has accepted a paper containing those results.

Fraud isn't unheard of in science. But because of peer review and the requirement of repeatability in science, bad research, whether the result of carelessness or fraud, is usually exposed quickly.

Steve and Dave Maslowski/Science Source

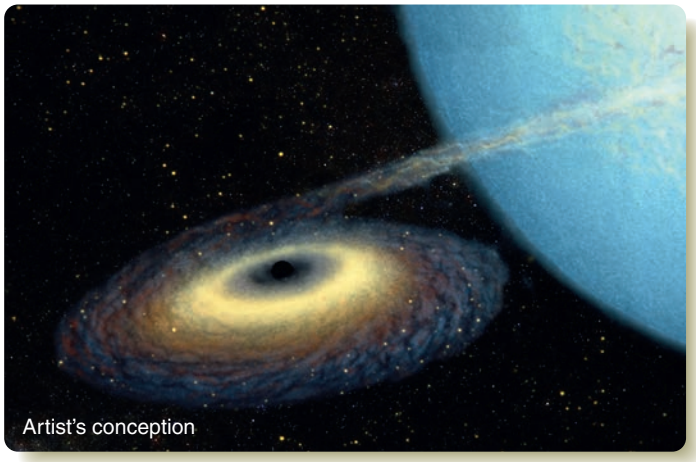
Chickadees always sing the same song: *Hey-sweetie-sweetie-sweetie.*

isolated black hole in space will not have much matter flowing into it, but a black hole in a binary system might receive a steady flow of matter transferred from its companion star. This suggests you can search for black holes by looking closely at X-ray binaries.

Some X-ray binaries such as Hercules X-1 contain a neutron star, and they will emit X-rays much as would a binary containing a black hole. You can tell the difference between an X-ray binary with a neutron star versus one with a black hole in two ways. If the compact object emits pulses, you know it is a neutron star because a neutron star has a solid surface on which sources of radiation can be anchored and cross into and out of Earth's view as the neutron star spins (Figure 3 of the Concept Art: The Lighthouse Model of Pulsars). With no solid surface, a black hole would not be able to emit an extended series of regular pulses. Another clue depends on the mass of the object. If the mass of the compact object is greater than about 3 $M_{\odot}$, the maximum mass that can be supported by degenerate neutrons, the object can't be a neutron star; it must be a black hole.

The first X-ray binary suspected of harboring a black hole was Cygnus X-1. It contains a supergiant B0 star and a compact object orbiting each other with a period of 5.6 days. Astronomers suspected that the observed X-rays were emitted by matter from the star flowing onto or into the compact object. The compact object is invisible, but Doppler shifts in the spectrum reveal the motion of the B0 star around the center of mass of the binary. From the geometry of the orbit, astronomers were able to calculate that the mass of the compact object had to be greater than 3.8 $M_{\odot}$, well above the maximum for a neutron star.

To confirm that black holes exist, astronomers needed a conclusive example—an object that couldn't be anything except a black hole. Cygnus X-1 didn't quite pass that test when it was first discovered. Perhaps the B0 star was not a normal star, and its portion of the system mass was incorrectly estimated; perhaps



Don Dixon/ cosmographica.com

▲ **Figure 14-14** The X-ray source Cygnus X-1 consists of a supergiant B0 star and a compact object orbiting each other. Gas from the B0 star's stellar wind flows into the hot accretion disk around the compact object, and the X-rays astronomers detect come from the disk.

the system contained a third star. Either possibility would distort the analysis.

It took years of work to understand Cygnus X-1. Further observations and analysis show that the B0 star has a mass of about 25 $M_\odot$, and the compact object has about 10 $M_\odot$. Astronomers conclude that matter flows from the B0 star as a strong stellar wind, and much of that matter goes across the L_1 point (look back to Figure 13-6) and ends up in a hot accretion disk about five times larger in diameter than the orbit of Earth's Moon. The inner few hundred kilometers of the disk have a temperature of about 2 million K—hot enough to radiate X-rays (**Figure 14-14**). The evidence is now strong that Cyg X-1 contains a black hole.

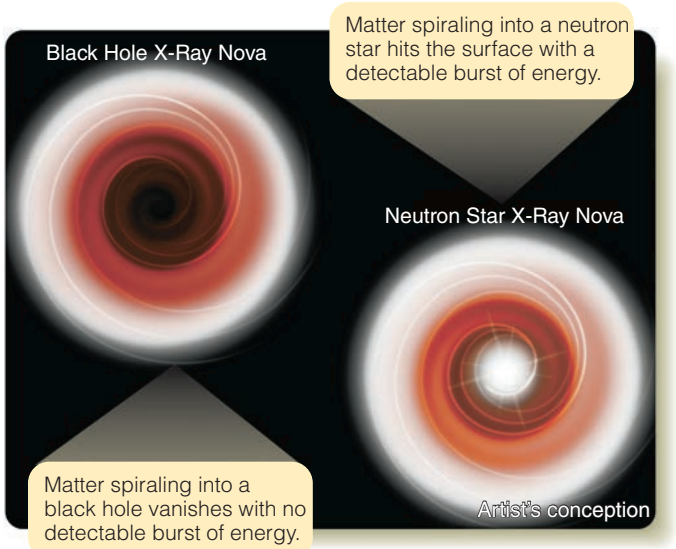
As X-ray telescopes have found many more X-ray–emitting objects, the list of binary systems containing black holes has grown. A selection of these objects is shown in **Table 14-2**. Each compact object is surrounded by a hot accretion disk in a close X-ray binary system. Some of the binary systems are easier to analyze than others, but it has become clear finally that these compact objects, including Cygnus X-1, are too massive to be neutron stars and must be black holes. And stellar remnant black holes almost certainly exist all over the Universe: Astronomers using the *Chandra X-ray Observatory* have identified 26 black hole candidates in the galaxy M31, which is a large neighbor to our Milky Way Galaxy.

Another way to confirm the identity of a black hole is to search for evidence of the defining characteristic—an event horizon—and that search also has been successful. In one study, astronomers selected 12 X-ray binary systems, six of which seemed to contain neutron stars and six of which were thought to contain black holes. Using X-ray telescopes, the astronomers monitored the systems, watching for telltale flares of energy as blobs of matter entered the accretion disks and spiraled inward. In the six systems thought to contain neutron stars, the astronomers could detect final bursts of energy when the blobs of matter finally hit the surfaces of the neutron stars. In the six systems suspected of containing black holes, however, the blobs of matter spiraled inward through the accretion disks and vanished without final bursts of energy. Evidently, those blobs of matter became undetectable as they approached the event horizons (**Figure 14-15**). This is dramatic evidence that event horizons are real.

The evidence shows that black holes really do exist. The next challenge is to understand how some of these objects interact with the matter flowing into them through accretion disks to produce high-energy jets and outbursts.

Table 14-2 Six Black Hole Binaries

Object	Star	Orbital Period	Mass of Black Hole
Cygnus X-1	B0 I	5.6 days	10 $M_\odot$
LMC X-3	B3 V	1.7 days	> 8 $M_\odot$
A0620-00	K6 V	7.75 hours	11 ± 1.9 $M_\odot$
V404 Cygni	K3 III	6.47 days	12 ± 3 $M_\odot$
GRO J1655-40	F5 IV	2.61 days	6.9 ± 1 $M_\odot$
QZ Vulpecula	K5 V	8.3 hours	10 ± 4 $M_\odot$


Black Hole X-Ray Nova
Matter spiraling into a neutron star hits the surface with a detectable burst of energy.
Neutron Star X-Ray Nova
Matter spiraling into a black hole vanishes with no detectable burst of energy.
Artist's conception

NASA/CXC/SAO

▲ **Figure 14-15** Gas spiraling into an accretion disk grows hot, and as it nears the central object, a strong gravitational redshift makes it appear redder and dimmer. Systems containing a neutron star emit bursts of energy when the gas hits the surface of the neutron star, but such bursts are not seen for systems containing black holes. In those systems, the matter vanishes as it approaches the event horizon. This is direct observational evidence of an event horizon around black holes.

Practicing SCIENCE

If relativistic effects slow time and prevent outside observers from seeing matter cross an event horizon, how can infalling matter disappear into a black hole without a trace?

This question is one that a scientist would ask to double-check the conclusion that some accretion disks are definitely around black holes, not neutron stars.

Astronomers observed flares when matter hit the surfaces of neutron stars, but observed no flares when matter fell into a black hole. Although time slows near the event horizon, remember the gravitational redshift. Hot matter flowing into a black hole can emit X-rays, but as the matter nears the event horizon, the gravitational redshift lengthens the wavelengths dramatically. The matter vanishes not because it crosses the event horizon, but because its photons are shifted to undetectably long wavelengths.

Now review another basic principle of compact object physics. ***Why does matter become hot as it falls into a black hole?***

14-3 Compact Objects with Disks and Jets

Matter flowing onto a neutron star or into a black hole forms an accretion disk and that can produce some surprising phenomena. Astronomers are working hard to understand these peculiar effects.

14-3a Jets of Energy from Compact Objects

Observations show that some compact objects are emitting jets of gas and radiation in opposite directions. These jets are similar to the bipolar outflows ejected by protostars, but are much more powerful. The X-ray images in Figure 14-4 and in Figure 3 of the Concept Art: The Lighthouse Model of Pulsars show that some young pulsars, including the Crab Nebula and Vela pulsars, are ejecting jets of highly excited gas. Systems containing black holes can also eject jets. For example, the binary system GRO J1655-40 contains a black hole and has been observed at radio wavelengths sporadically ejecting jets in two opposite directions at 92 percent of the speed of light.

One of the most powerful examples of this process is an X-ray binary called SS 433. Its optical spectrum shows sets of spectral lines that are Doppler shifted by about one-fourth the speed of light, with one set shifted to the red and one set shifted to the blue. Astronomers recognize the combination of red and blue Doppler shifts as evidence of oppositely directed jets. Furthermore, the two sets of spectral lines shift back and forth across each other with a period of 164 days.

Apparently, SS 433 is a binary system in which a compact object (probably a black hole) pulls matter from its companion star and forms an extremely hot accretion disk. Jets of high-temperature gas blast away from the disk in beams aimed in opposite directions (**Figure 14-16**). As the disk precesses, it sweeps these beams around the sky once every 164 days, and telescopes on Earth detect light from gas carried outward in both beams. One beam produces a redshift, and the other produces a blueshift.

The fact that jets are observed to contain intact atomic nuclei rather than electrons and positrons is evidence that the ultimate source of the jet material is the accretion disk, not matter and antimatter produced in the polar regions of the compact object. Somehow the extremely hot gas in an accretion

© 2005, Fahad Sulehria, www.novacelestia.com. Reprinted with permission.

▲ **Figure 14-16** In this artist's conception, matter from a normal star flows into an accretion disk around a compact object. Processes in the spinning disk eject gas and radiation in jets perpendicular to the disk.

disk can emit powerful beams of gas and radiation along the disk's axis of rotation (look again at Figure 14-16), The exact process isn't well understood, but it seems to involve magnetic fields that get caught in the accretion disk and are twisted into tightly wound tubes that squirt gas and radiation out of the disk and confine them in narrow jets and beams. However, it's not clear why some accretion disks produce jets and others don't.

Accretion disk jets are spectacular in their own right, but they are also hypothesized to be a mechanism, in addition to supernova shock waves, of accelerating protons and other atomic nuclei to extremely high energies. These particles travel at nearly the speed of light and reach Earth as cosmic rays (look back to Section 6-6a).

Pairs of jets from stellar remnants are the prototype of a phenomenon in which an accretion disk around a compact object produces powerful beams of radiation and matter that spray cosmic-ray particles across space. You will encounter this phenomenon again in a later chapter when you study active galaxies.

14-3b Gamma-Ray Bursts

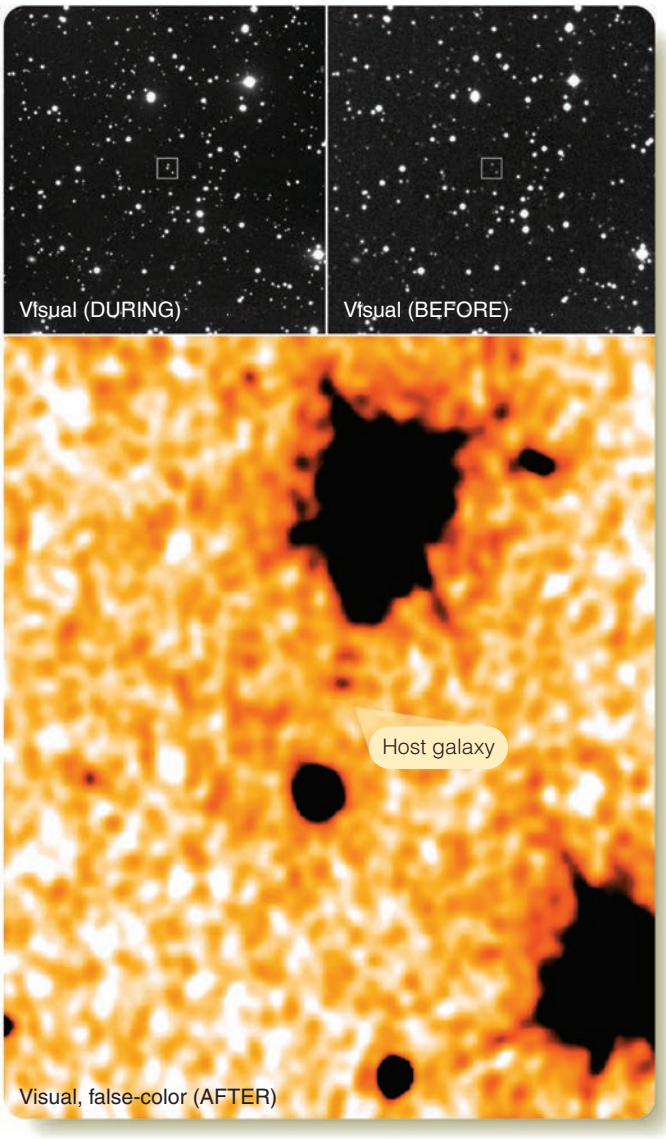
The Cold War played a minor part in the story of neutron stars and black holes. In 1963, the United States and the Soviet Union signed a nuclear test ban, and by 1970, the United States had finished launching a series of 12 satellites to watch for nuclear tests that were violations of the treaty. A nuclear detonation emits gamma-rays, so the satellites were designed to watch for bursts of gamma-rays coming from Earth. The experts were startled when the satellites detected **gamma-ray bursts (GRBs)** coming from space at the rate of about one per day. When those data were finally declassified in 1973, astronomers realized that the bursts might be coming from neutron stars and black holes.

The *Compton Gamma Ray Observatory (CGRO)* reached orbit in 1991 and immediately began detecting several GRBs per day. Its observations showed that the intensity of the gamma-rays rises to a maximum in seconds and then fades away quickly; a burst is usually over in a few seconds, almost always in less than a minute.

Data from the *CGRO* also showed that the GRBs were coming from all over the sky and not from any particular region. This helped astronomers eliminate some hypotheses. For example, initially there was a hypothesis that the GRBs were being produced by relatively common events involving the stars in our Galaxy, but the *CGRO* data eliminated that possibility. If the GRBs were produced among stars, you would expect to see them most often along the Milky Way where there are lots of stars. The fact that bursts occur all over the sky means that they must be produced by rare events in distant galaxies.

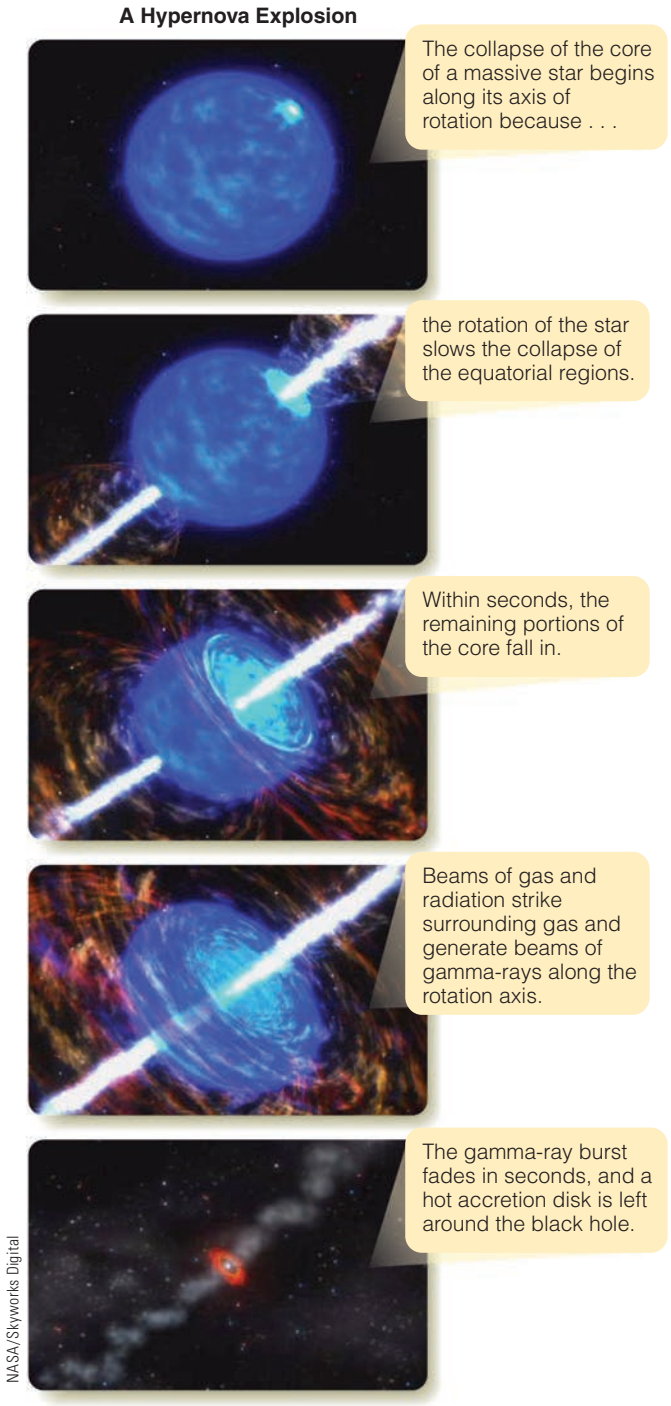
GRBs are hard to study because they occur without warning and fade so quickly, but starting in 1997 new satellites were put into orbit that were designed to overcome those obstacles. Their data show that there are two kinds of GRBs: Short bursts last less than 2 seconds, but longer bursts can go on for many seconds or even minutes. Specialized space observatories now can detect bursts, quickly determine their location in the sky, and immediately alert astronomers on the ground. When telescopes on Earth swivel to image the locations of the bursts, they detect fading glows that resembled supernovae (**Figure 14-17**), suggesting that long GRBs are produced by a certain kind of supernova explosion.

Stellar models of stars indicate that a star more massive than about 20 $M_{\odot}$ can exhaust its nuclear fuel and collapse directly into a black hole. Models show that if the collapsing star is initially rotating relatively rapidly, conservation of angular momentum would greatly increase that spin, slowing the collapse of the



▲ **Figure 14-17** Alerted by gamma-ray detectors on satellites, observers used one of the VLT 8.2-m telescopes on a mountaintop in Chile to image the location of a GRB only hours after the burst. The image at top left shows the fading glow of the eruption (*box*). The image at top right, recorded 13 years before, reveals no trace of an object at the location of the GRB. The *Hubble Space Telescope* image at bottom was recorded a year later and reveals a very faint, distant galaxy at the location of the GRB.

equatorial parts of the star. The poles of the star would fall in more quickly, and that would focus beams of intense radiation and ejected gas that would blast out along the axis of rotation. Such an eruption is called a **hypernova** (Figure 14-18). If one of those beams were pointed at Earth, it would appear as a powerful GRB. The long GRBs are probably produced by hypernovae.

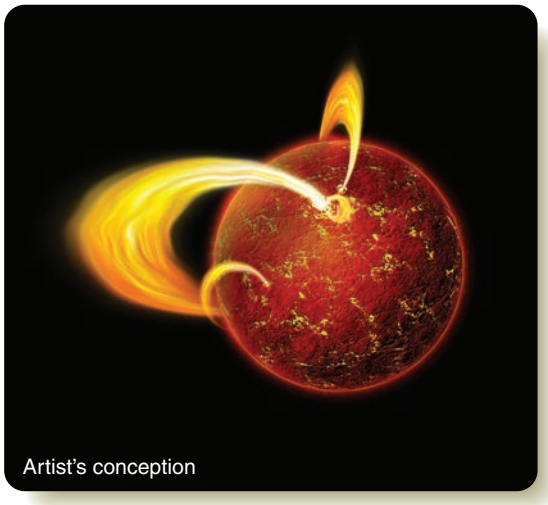

A Hypernova Explosion
The collapse of the core of a massive star begins along its axis of rotation because . . .
the rotation of the star slows the collapse of the equatorial regions.
Within seconds, the remaining portions of the core fall in.
Beams of gas and radiation strike surrounding gas and generate beams of gamma-rays along the rotation axis.
The gamma-ray burst fades in seconds, and a hot accretion disk is left around the black hole.
NASA/Skyworks Digital

▲ **Figure 14-18** Artist's Conception: When an extremely massive star collapses in a hypernova explosion, energy from the core is focused into beams of radiation and matter that point along the axis of rotation. This is thought to be the source of GRBs that last longer than 2 seconds.

In 2008, the *Swift* orbiting telescope detected an intense GRB that originated in a galaxy 7.5 billion ly from Earth. The burst was so powerful that, for about one minute, its visual-wavelength component was bright enough to see with the unaided eye. If you had been looking directly at it, you would have seen it appear as a "nova" star slightly brighter than those in the Little Dipper. Another powerful GRB detected by *Swift* in 2008 originated in a galaxy 12.2 billion ly away. Despite the distance, it was one of the brightest GRBs ever detected. Astronomers suspect that both of these bursts were produced by hypernova collapses of massive stars in which one of the jets was aimed directly at Earth.

Not all long GRBs produce visible afterglows, and at first, astronomers thought they were a different kind of eruption. But X-ray telescopes have observed some of these "dark" GRBs and detected afterglows at X-ray wavelengths. This suggests that some long GRBs are obscured by dust so that their afterglows are visible only at X-ray wavelengths. The dark GRBs probably are not a new kind of GRB.

Short GRBs are different and don't seem to be associated with hypernovae. Some repeat, and these repeating bursts seem to be produced by neutron stars with magnetic fields 100 times stronger than that in a normal neutron star. Dubbed **magnetars**, these objects can produce bursts of gamma-rays when shifts in the magnetic field break the crust of the neutron stars, causing "starquakes" that release large amounts of energy (Figure 14-19). The *Fermi* observatory detected a neutron star that has gamma-ray eruptions as often as 100 times in 20 minutes. Another magnetar produced a burst of gamma-rays that reached Earth in 1998 and was strong enough to significantly increase the ionization of Earth's upper atmosphere, momentarily disrupting radio communication worldwide. Changes detected in the Cassiopeia A



▲ **Figure 14-19** Some neutron stars appear to have magnetic fields up to 1000 times stronger than those in a normal neutron star. These magnetars can produce bursts of gamma-rays when shifts in the magnetic field rupture the rigid crust of the neutron star.

supernova remnant (Figure 13-18b) appear to have been caused by an eruption on the central neutron star that occurred about 1953. Eruptions on the surfaces of magnetars are so powerful they evidently can produce noticeable effects many light-years away.

A model has been proposed in which all neutron stars begin their lives as magnetars. Then, as their magnetic fields gradually decrease in strength, they evolve into classic radio pulsars. Finally, they become a type of radio-quiet but X-ray–bright neutron star, of which only a few have been observed.

There is evidence that not all short GRBs are produced by magnetars. Some high-energy bursts have occurred in parts of galaxies far from star-forming regions, places where you would not expect to find the young, massive stars that produce magnetars or hypernovae. Also, the afterglows don't resemble fading supernovae, and these bursts do not repeat. Such bursts may be produced by the merger of two neutron stars that orbit each other, radiate orbital energy as gravitational radiation, and eventually spiral into each other. Such a collision would cause a violent explosion as the two objects merged to form a black hole. Some model calculations indicate that neutron star mergers may actually produce some heavy elements such as gold in amounts comparable to supernova explosions.

Other short GRBs are probably produced by the merger of a neutron star with a black hole. As these objects spiral into each other, the neutron star would be ripped apart by tidal forces before being swallowed by the black hole. Model calculations indicate that the GRB and afterglow produced by a neutron star being swallowed by a black hole should differ from the result of the merger of two neutron stars. Astronomers are now working to distinguish between these two kinds of short GRBs.

Could a GRB occur near Earth? The nearest known binary pulsar is only about 2000 ly from Earth. If a GRB occurred at that distance, the gamma-rays would shower Earth with radiation equivalent to a 10,000-megaton nuclear blast, comparable to a full-scale nuclear war between superpowers. (The largest single weapon ever detonated released less than 60 megatons of energy.) The gamma-rays could create enough nitric oxide in the atmosphere to produce intense acid rain and also would destroy the ozone layer, exposing life on Earth to deadly levels of solar ultraviolet radiation. Even if the GRB occurs far across our Galaxy, Earth might be affected if the gamma-ray beam is pointed right at us. It is possible that Earth could be affected by a GRB as often as once every million years. Such events could be among the causes of the mass extinctions that show up in the fossil record.

Does it surprise you that such astonishing events as merging neutron stars and hypernovae produce something so common that gamma-ray telescopes observe one or more every day? Remember that these events are so powerful they can be detected over very great distances. There may be 30,000 neutron star binaries in each galaxy, and there are hundreds of billions of galaxies within range of gamma-ray telescopes. Earthlings are treated to the entire observable Universe's display of these cosmic catastrophes.

WHAT ARE WE?

Unexciting

Look around. What do you see? A table, a chair, a tree? It's all simple, ordinary, unexciting stuff. The world we live in is familiar and comfortable, but astronomy reveals that much of the Universe is utterly unlike anything you have ever experienced.

Throughout the Universe, gravity makes clouds of gas form stars, and in turn, the stars generate energy through nuclear fusion in their cores, which delays gravity's final victory. But gravity always wins. You have learned that stars of different masses die in different ways, but you have also discovered that they always reach one of three end states: white dwarf, neutron star, or black hole. However strange these compact objects may seem in Earthly terms, there are billions of them in the Universe. They are common.

The physics of compact objects is extreme and violent. You are not accustomed to objects as hot as the surface of a neutron star, and you have never experienced the environment near a black hole, where gravitational tidal forces are so strong they would pull you to pieces.

The Universe is filled with things that are so violent and so peculiar they are almost unimaginable, but such things can almost be called common. Next time you are out for a walk, look around and notice how beautiful and peaceful Earth can be and ponder how unusual that may be compared with the rest of the Universe.

Study and Review

SUMMARY

14-1 Neutron Stars

- After a type II supernova, theory predicts that if the star core's mass is between 1.4 $M_\odot$ and about 3 $M_\odot$, the core will halt its contraction as a **neutron star**. A neutron star is supported by neutron degeneracy pressure. Theory predicts that a 1.4-$M_\odot$ neutron star should be about 10 km in radius, spin very fast, and have a powerful magnetic field.
- **Pulsars**, which emit precisely timed short radio bursts, were discovered in 1967. These short bursts and the discovery of a pulsar in the supernova remnant called the Crab Nebula were key evidence that pulsars are neutron stars.
- The **lighthouse model** explains pulsars as spinning neutron stars that emit radiation beams from their magnetic poles. As pulsars spin, they sweep the beams around the sky like lighthouses; if the beams sweep over Earth, astronomers detect these radio pulses.
- Most of the energy emitted by a pulsar is carried away by a **pulsar wind**. Some energy is carried away by the beams of radiation emitted by young neutron stars, ultimately driven by their rapid rotation. A spinning neutron star slows as it ages, thus decreasing the amount of energy it radiates into space via radiation beams.
- Many pulsars have been found in binary systems. Observations of the first binary discovered containing two neutron stars revealed that the system is losing orbital energy by emitting **gravitational radiation**. **X-ray bursters** are systems in which matter from a companion star accumulates on the surface of the neutron star until the matter undergoes nuclear fusion and explodes off the surface of the neutron star, after which the cycle repeats.
- The fastest pulsars are **millisecond pulsars**, which emit more than 1000 radio pulses per second. Millisecond pulsars appear to be systems in which an old neutron star has been spun up to high speeds by mass flowing from its binary companion.
- Extrasolar planets have been found orbiting at least two neutron stars. They may be the remains of a companion star that was mostly devoured by the neutron star, or they may have formed from a ring of gas and dust left after the supernova explosion that created the neutron star.

14-2 Black Holes

- If the core of a type II supernova has a mass greater than 3 $M_\odot$, then it will contract to a very small size—perhaps to a **singularity** with zero radius. Near such an object, gravity is so strong that not even light can escape and that region is called a **black hole**. The outer boundary of the volume occupied by a black hole is the **event horizon.** The radius to the event horizon is the **Schwarzschild radius, R_S**, which amounts to only 3 kilometers for a black hole of 1 $M_\odot$.

$$R_s = \frac{2GM}{c^2} \qquad \text{(Eq. 14-1)}$$

- If you were to leap into a black hole with a ticking clock and a lit flashlight or torch, your friends who stayed behind would see two relativistic effects. They would see your clock slow relative to their own clock because of **time dilation.** Also, they would see your light redshifted to longer wavelengths because of the **gravitational redshift**. You would not notice these effects; your clock would tick once per second and your flashlight would emit white light. However, you would feel powerful tidal forces that would deform and heat your body until you grew hot enough to emit X-rays. Any X-rays you emitted before reaching the event horizon could escape.
- To search for black holes, astronomers must look for binary star systems in which mass flows onto a compact object and emits X-rays. If the mass of a compact object is clearly greater than about 3 $M_\odot$, then the object is presumably a black hole. Clues that an X-ray source is caused by accretion onto a black hole rather than onto a neutron star are a lack of sustained regular pulsations and a lack of flares when blobs of material in the accretion disk arrive at the central compact object.

14-3 Compact Objects with Disks and Jets

- Powerful jets of gas and radiation have been detected being emitted perpendicular to accretion disks around neutron stars and black holes
- **Gamma-ray bursts (GRBs)** appear to be related to violent events involving neutron stars or black holes. Bursts longer than 2 seconds appear to arise during **hypernovae** caused by rapidly rotating stars more massive than 20 $M_\odot$ collapsing to become black holes. Some short GRBs that repeat may be produced by shifts in the powerful magnetic fields in a type of pulsar called **magnetars**. Some other short GRBs that do not repeat may be from the merger of binary compact objects such as neutron-star pairs or neutron-star/black-hole pairs.

REVIEW QUESTIONS

1. How are neutron stars and white dwarfs similar? How do they differ?
2. Compare a 1.4-$M_\odot$ isolated neutron star with a 3.0-$M_\odot$ isolated neutron star. Which would you expect to rotate faster? Which would have the smaller magnetic field? Which would be smaller in radius? Why?
3. Consider an isolated 12-$M_\odot$ main-sequence star. Which type of compact object will it evolve into: a white dwarf, a neutron star, or a black hole? How do you know? Are you sure? Why or why not?
4. Consider an isolated 5-$M_\odot$ main-sequence star. Which type of compact object will it evolve into: a white dwarf, a neutron star, or a black hole? How do you know? Are you sure? Why or why not?
5. Consider an isolated 25-$M_\odot$ main-sequence star. Which type of compact object will it evolve into: a white dwarf, a neutron star, or a black hole? How do you know? Are you sure? Why or why not?
6. Why do neutron stars have an upper mass limit?
7. Why do you expect neutron stars to spin more rapidly than white dwarfs?
8. If neutron stars have hot surface temperatures, why aren't they very luminous?
9. Where would you put neutron stars on the H–R diagram? Assume the surface temperature of a neutron star is around 1 million K.
10. Why do you expect neutron stars to have a more powerful magnetic field than a white dwarf?

11. You would like to observe a neutron star that has an apparent visual magnitude of +21. Which kind of telescope would you need to observe this object: a backyard telescope you bought from a local big-box store, a small observatory on the campus of a university in a large city, the Gran Telescopio Canarias, or the *Hubble Space Telescope*? (*Hints:* See Figure 2-6, plus Sections 6-3 and 6-4.)
12. Do pulsars change in size as they emit radio bursts? How do you know?
13. How did astronomers conclude that pulsars actually could not be pulsating stars?
14. Why would astronomers naturally assume that the first discovered millisecond pulsar was relatively young?
15. How can a neutron star in a binary system generate X-rays?
16. Compared to the speed of light, what velocity would your spaceship need to escape a neutron star of 3 $M_\odot$ and radius 10 km?
17. If the Sun has a Schwarzschild radius, why isn't it a black hole?
18. How can a black hole emit X-rays?
19. Are black holes degenerate objects, meaning, are they supported by a degeneracy pressure?
20. What is the inferred minimum mass to an isolated black hole?
21. Can the Sun ever become a black hole? Why or why not?
22. In what sense is a black hole actually black?
23. If you are falling into a black hole and you point the white light from your flashlight away from the black hole, would the wavelengths of photons from the flashlight received by a distant observer shift toward the red or the blue end of the electromagnetic spectrum, or neither?
24. What evidence can you cite that black holes really exist?
25. How can mass transfer into a compact object produce jets of high-speed gas? X-ray bursts? GRBs?
26. Discuss some possible causes of short GRBs; of long GRBs.
27. **How Do We Know?** How does peer review make fraud rare in science?

ACTIVE INQUIRY QUESTIONS

1. You are reading a novel about an advanced alien civilization on a planet that orbits a neutron star. From what you know, what can you say about the likelihood of such a situation?
2. You are on a planet that orbits Vega. You observe a pulsar. Would a person on Earth necessarily also see the same star as a pulsar? Why or why not?
3. An eccentric philanthropic billionaire is willing to fund your team in a search for undiscovered black holes. Where do you look, and what kind of instrument(s) do you need to acquire or build to perform the search?

PROBLEMS

1. Because of the properties of degenerate matter, (*Note:* a year is 3.2×10^7 s.) white dwarfs follow a mass–radius relationship: R is proportional to $M^{-1/3}$, where R is radius and M is mass. If a 0.75-$M_\odot$ white dwarf has a radius of 7000 km, what is the radius of a 1.25-$M_\odot$ white dwarf?
2. Neutron stars are composed of degenerate matter and therefore follow a mass–radius relationship of the same form as for white dwarfs: R is proportional to $M^{-1/3}$, although with nearly 1000 times smaller size for the same mass. If a 1.4-$M_\odot$ neutron star has a radius of 10 km, what is the radius of a 3.0-$M_\odot$ neutron star?
3. What is the density of a 1.0-$M_\odot$ white dwarf? A 1.4-$M_\odot$ white dwarf? Some necessary data are provided in Problem 1. (*Notes:* Density is mass divided by volume; the volume of a sphere is $\frac{4}{3}\pi r^3$; the mass of the Sun can be found in Celestial Profile: The Sun, Chapter 8.)
4. What is the density of a 1.4-$M_\odot$ neutron star? A 3.0-$M_\odot$ neutron star? Necessary data are provided in Problem 2. (*Notes:* Density is mass divided by volume; the volume of a sphere is $\frac{4}{3}\pi r^3$; the mass of the Sun can be found in Celestial Profile: The Sun, Chapter 8.)
5. What is the escape velocity from the surface of a 1.0-$M_\odot$ white dwarf? From a 1.4-$M_\odot$ white dwarf? Some necessary data are provided in Problem 1. (*Hints:* Use the formula for escape velocity, Eq. 5-1b; make sure to express quantities in units of meters, kilograms, and seconds.)
6. What is the escape velocity from the surface of a 1.4-$M_\odot$ neutron star? From a 3.0-$M_\odot$ neutron star? Necessary data are provided in Problem 2. (*Hints:* Use the formula for escape velocity, Eq. 5-1b; make sure to express quantities in units of meters, kilograms, and seconds.)
7. If a neutron star has a radius of 10 km and a temperature of 1.0×10^6 K, how luminous is it? Express your answer in watts and also in solar luminosity units. (*Note:* the luminosity of the Sun can be found in Celestial Profile: The Sun, Chapter 8.) (*Hint:* Use the Stefan-Boltzmann law, Eq. 7-1 and refer to Figure 6-3.)
8. What is the peak wavelength of a neutron star's luminosity? Necessary data are provided in Problem 7. Express your answer in units of nm. In which band of the electromagnetic spectrum is that wavelength? (*Hint:* Use Wien's law, Eq. 7-2.)
9. If a neutron star has a radius of 10 km and rotates 716 times a second, what is the speed at which a point on the surface at the neutron star's equator is moving? Express your answer as a fraction of the speed of light. (*Note:* The speed of light is 3×10^5 km/s.)
10. What is the separation distance of the binary pulsar PSR 1913+16 if both neutron stars have about 1.4 $M_\odot$? Does this value match that discussed in the text? Why or why not? Some necessary data can be found in the chapter text. (*Hints:* Use the version of Kepler's third law for binary stars, Eq. 9-4; make sure to express quantities in units of AU, solar masses, and years.)
11. A 1.4-$M_\odot$ neutron star and a 1.0-$M_\odot$ white dwarf have been found orbiting each other with a period of 11 minutes. What is their average separation? Convert your answer to units of the Sun's radius, which is 0.0047 AU. (*Hints:* Use the version of Kepler's third law for binary stars, Eq. 9-4; make sure to express quantities in units of AU, solar masses, and years.) (*Note:* a year is 3.2×10^7 s.)
12. If a circular accretion disk around a 1.4-$M_\odot$ neutron star has a radius of 2.0×10^5 km as measured from the center of the neutron star to the edge of the disk, what is the orbital velocity of a gas particle located at its outer edge? (*Note:* The mass of the Sun can be found in Celestial Profile: The Sun, Chapter 8.) (*Hints:* Use the formula for circular velocity, Eq. 5-1a; make sure to express quantities in units of meters, kilograms, and seconds.)
13. If Earth's Moon were replaced with a typical neutron star, what would the angular diameter of the neutron star be as seen from Earth? (*Note:* The distance between Earth and the Moon can be found in Appendix Table A-11.) (*Hint:* Use the small-angle formula, Eq. 3-1.)
14. What is the escape velocity at the Schwarzschild radius of a 3-$M_\odot$ black hole? A 10-$M_\odot$ black hole? Some necessary data can be found in the chapter text. (*Note:* The mass of the Sun can be found in Celestial Profile: The Sun, Chapter 8.) (*Hint:* Use the escape velocity formula, Eq. 5-1b.)

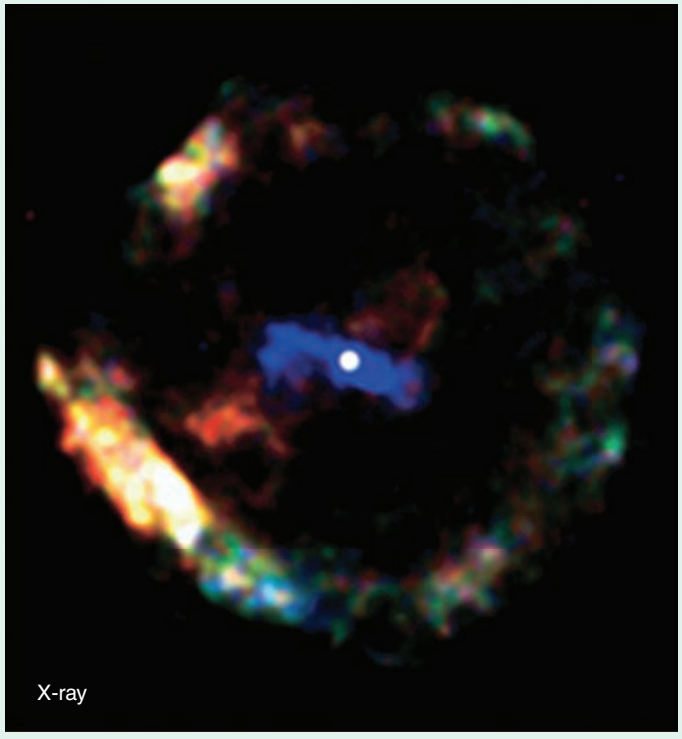
15. What is the orbital period of a bit of matter in an accretion disk that is located 2.0×10^5 km from the center of a 10-$M_\odot$ black hole? (*Note:* The mass of the Sun can be found in Celestial Profile: The Sun, Chapter 8.) (*Hints:* Use the circular velocity formula, Eq. 5-1a; make sure to express quantities in units of meters, kilograms, and seconds.) (*Note:* a year is 3.2×10^7 s.)
16. If an X-ray binary consists of a 20-$M_\odot$ star and a 1.4-$M_\odot$ neutron star orbiting each other every 13.1 days, what is their average separation? (*Hints:* Use the version of Kepler's third law for binary stars, Eq. 9-4; make sure to express quantities in units of AU, solar masses, and years.)

SENSE OF PROPORTION

1. Arrange the following in order of increasing size: the Sun; a 1.4-$M_\odot$ white dwarf; a 3-$M_\odot$ neutron star; a 3-$M_\odot$ black hole.
2. Earth is about 650 times larger in diameter than a typical neutron star. If you represent a neutron star with a golf ball (diameter 1.68 in.), how big a ball would you need to represent Earth?

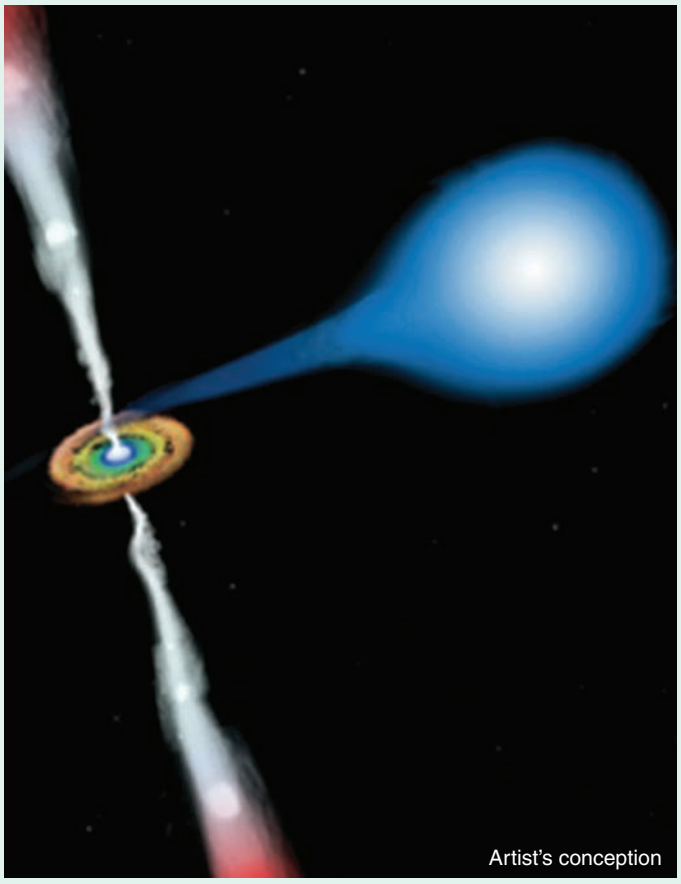
LEARNING TO LOOK

1. This X-ray image shows the supernova remnant G11.2-0.3 and its central pulsar in X-rays. The blue nebula near the pulsar is caused by the pulsar wind. How old do you think this system is? Discuss what the appearance of this system might be a million years from now.



NASA/CXC/SAO/Eureka Scientific/M. Roberts et al.

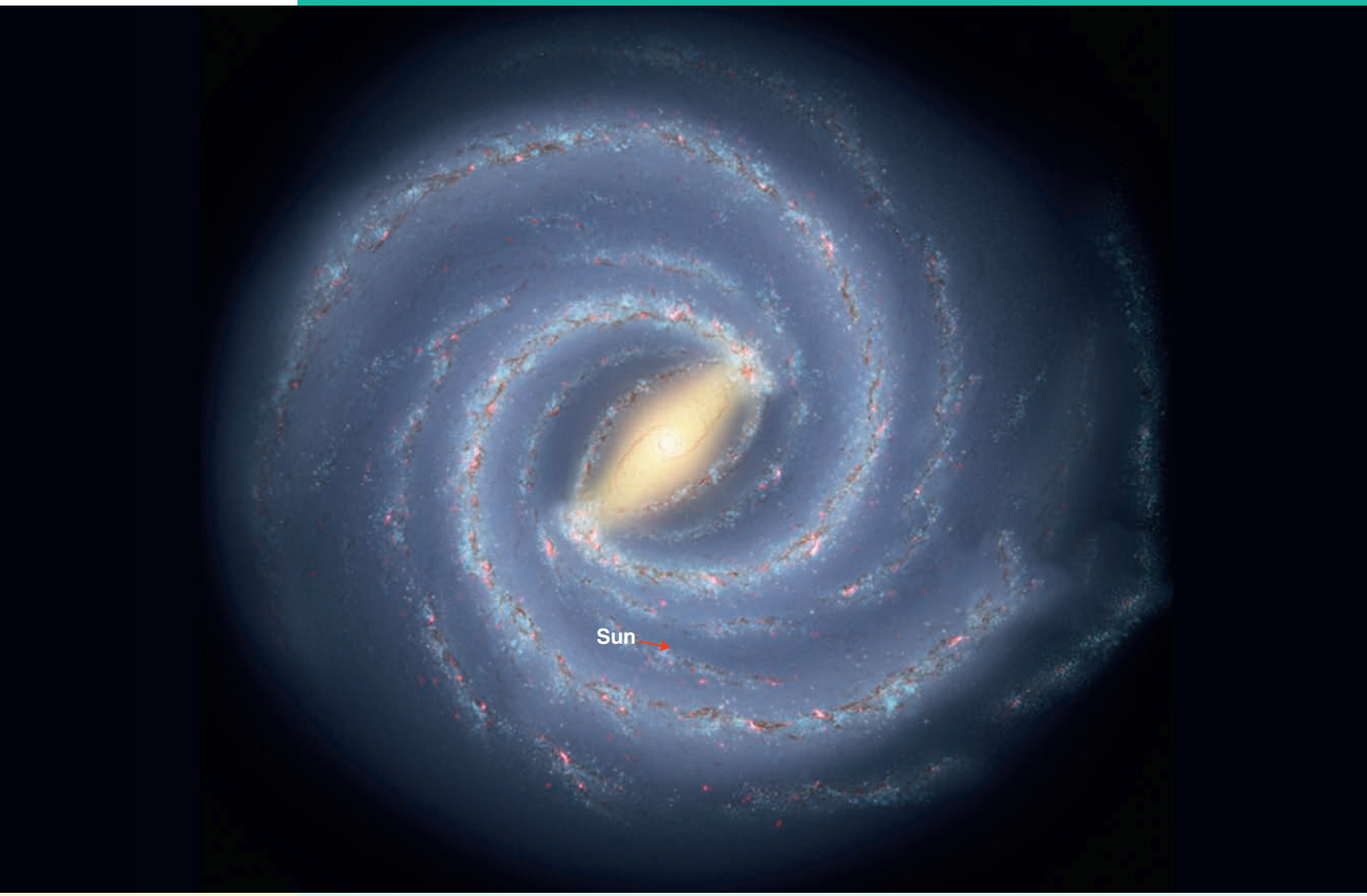
2. What is happening in the artist's conception shown here? How would you distinguish between a neutron star and a black hole in such a system?



NASA/CXC/SAO/M. Weiss

3. Look at Figure 14-3. When would you expect to see the next main pulse and the next secondary pulse occurring in visual wavelengths? Express your answers in milliseconds.
4. Compare the lighthouse model with the visual-wavelength image of the Crab pulsar in the Concept Art: The Lighthouse Model of Pulsars. List and explain the differences.
5. Look at Figure 14-7. Where is Earth located in Figure 14-7b if we see the X-rays turn off and on, as in Figure 14-7a?

15 The Milky Way Galaxy



NASA/JPL-Caltech/Robert Hurt

▲ Artist's conception of the Milky Way Galaxy as seen face-on from outside, based on data from ground-based and space observatories. The location of our Sun, Earth, and Solar System is indicated about halfway from the center to the lower edge of the Galaxy. Notice the large central bar.

You have traced the life stories of stars from their birth in clouds of gas and dust to their deaths as white dwarfs, neutron stars, or black holes. Now you are ready to step back and view the vast groups of stars called galaxies. This chapter focuses on our home galaxy, the Milky Way, and five important questions:

- What is the evidence that we live in a galaxy?
- What is the evidence that our Milky Way Galaxy is a spiral galaxy?
- What are the spiral arms of the Milky Way and other spiral galaxies?
- What is in the nucleus of the Milky Way Galaxy?
- How did the Milky Way Galaxy form and evolve?

Discovering the Milky Way Galaxy by answering these questions will be one more step toward understanding the Universe as a whole. In the chapters that follow, you will leave our home Galaxy and voyage out among the billions of other galaxies that fill the depths of space.

The Stars Are Yours

JAMES PICKERING

THE STARS ARE YOURS is the title of a popular astronomy book written by James Pickering in 1948. The title expresses the author's view that the stars belong to everyone, equally, and you can enjoy the stars as if you owned them.

Next time you admire the night sky, recall that every star you see is part of the star system in which you live. You will learn in this chapter how evidence reveals that we are inside a great wheel of stars, a galaxy. The Milky Way Galaxy is more than 80,000 ly in diameter and contains more than 100 billion stars. It is our Galaxy because we live in it, but we are also products of it because the stars in the Milky Way Galaxy made most of the atoms in Earth and in our bodies. You can begin this chapter by pondering the notion that the stars belong to you, but when you reach the end of the chapter, you might realize that you also belong to the stars.

15-1 Discovery of the Galaxy

It seems odd to say that astronomers discovered something that is all around us, but it isn't obvious that we live in a galaxy. You might ask, "How do we know what our Galaxy is like? Nobody has ever seen it from the outside." Finding the evidence to answer that question was one of the greatest adventures in astronomy.

15-1a The Great Star System

Since ancient times, humanity has been aware of a hazy band of light around the sky (**Figure 15-1**). The ancient Greeks named that band *galaxies kuklos,* the "milky circle." The Romans changed the name to *via lactia,* meaning "milky road" or "milky way." It was not until Galileo used his telescope in 1610 that anyone realized the Milky Way is made of a huge number of stars.

Almost every celestial object you can see with your unaided eyes is part of the Milky Way Galaxy. The only exception viewable from Earth's Northern Hemisphere is the Andromeda Galaxy (also known as Messier 31), which is barely visible as a faint patch of light in the constellation Andromeda. You can use the star charts in Appendix B to locate the Milky Way and the Andromeda Galaxy (labeled M31). Later in this chapter you will learn that our Galaxy is also a spiral galaxy and, seen from a distance, would look like the artist's conception in the image that opens this chapter.

Galileo's telescope revealed that the glowing Milky Way is made up of millions of faint stars. Later astronomers realized that the shape of the Milky Way means that the Sun and Earth must be located inside a great wheel-shaped cloud of stars that they called the "star system." The band of the Milky Way encircling the sky is how that star system appears from our location inside.

ESO/Y. Beletsky

◀ **Figure 15-1** Nearby stars look bright, and people around the world group them into different constellations. Nevertheless, the vast majority of the stars in our Galaxy that are visible from Earth merge into a faintly luminous path that circles the sky: the Milky Way. This artwork shows the location of a portion of the Milky Way near a few bright winter constellations. (See Figure 15-3 and the star charts in Appendix B to locate other sections of the Milky Way in your sky.)

In 1750, Thomas Wright, drawing on the technology of the time, referred to the wheel-shaped star system as the "grindstone universe," by analogy with the thick disks of stone used in mills. Wright used the term *universe* because, so far as was known at the time, the Milky Way star system was the entire Universe.

In the late 18th century, astronomers Sir William Herschel and Caroline Herschel (Sir William's sister) set out to map the three-dimensional shape of the Milky Way. They assumed that they could see to the outer boundaries of the Milky Way in all directions and hypothesized that by counting the number of stars that were visible in different directions they could find the relative distances to its edges. If they saw many stars in one direction, they reasoned that the edge of the Milky Way in that direction is far away. If their telescope revealed fewer stars in another direction, they concluded that edge must be closer. Calling their method "star gauging," they counted stars in 683 directions in the sky and outlined a model of the star system (**Figure 15-2**). Their data (incorrectly) indicated that the Sun is near the center of the disk-shaped star system. In some directions, the Herschels saw very few stars, and these "holes in the sky" produced great irregularities along the edge of their diagram.

The model proposed by the Herschels was widely accepted and studied by other astronomers. The Herschels were not able to measure the size of the star system, but later researchers attempted to do so. By the early 20th century, astronomers had incorrectly concluded that the star system's disk is about 10 kiloparsecs in diameter and 2 kiloparsecs thick. A **kiloparsec (kpc)** is 1000 parsecs, about 3300 ly.

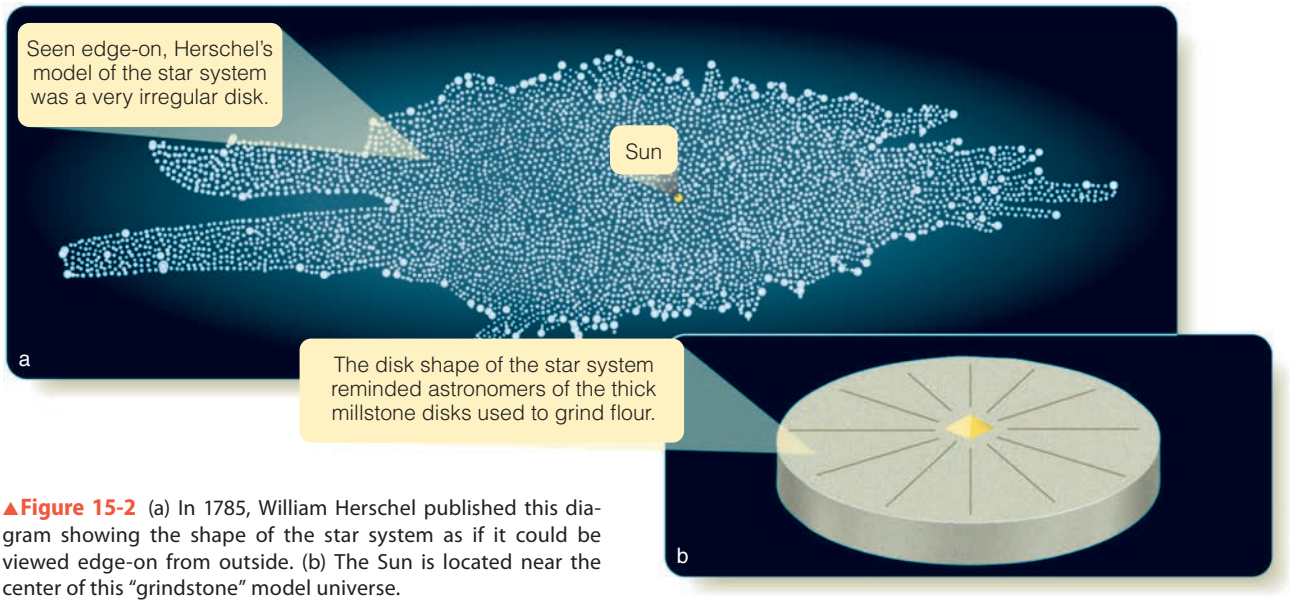
The Herschels assumed that they could see to the edge of the star system, but that was not correct. They were actually seeing only as far into the Milky Way as allowed by interstellar dust. At the time the Herschels did their work, astronomers did not understand that the interstellar medium partially blocks the passage of light (look back to Section 10-1b). Because the Herschels counted approximately the same numbers of stars in most directions around the Milky Way, they incorrectly concluded that the Sun is near the center of the star system. Furthermore, the holes in the sky observed by the Herschels are not empty but are now known to be especially dense interstellar clouds completely blocking the view of stars beyond them.

Modern astronomers know that the Milky Way Galaxy is much larger than was first thought and that the Sun is not at its center. How the human race realized the truth about our location in the Galaxy is a story that unfolds in the next section.

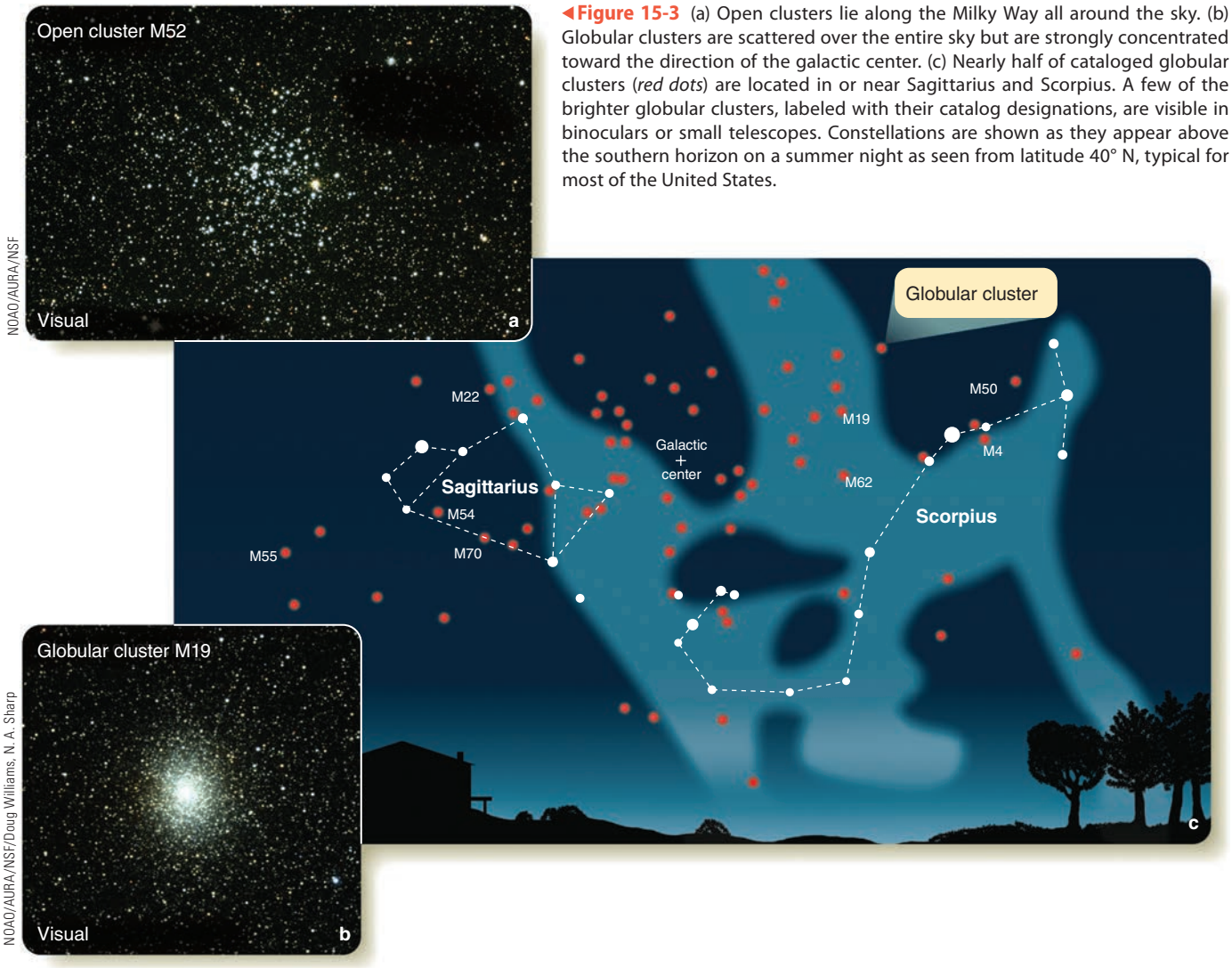
15-1b Variable Stars and the Size of the Galaxy

Humanity's understanding of the nature of the Milky Way began to change radically around the year 1920 when an astronomer named Harlow Shapley discovered how big the star system really is. Besides being one of the turning points of modern astronomy, Shapley's study illustrates one of the most common techniques in astronomy. If you want to understand some of the basic ways astronomers find out about the Universe, then Shapley's part of the story is worth following in detail.

You learned in Section 12-3a that there are two quite different types of star clusters, open clusters and globular clusters. Shapley began his research on the Milky Way star system by noticing that, although open star clusters are scattered all along the plane of the Milky Way, more than half of all globular clusters lie in or near the constellation Sagittarius. The globular clusters seem to be distributed in a great cloud with a center located somewhere off in that direction (**Figure 15-3**). Shapley assumed that the orbital motion of these clusters is controlled by the gravitation of the entire star system, and for that reason he


Seen edge-on, Herschel's model of the star system was a very irregular disk.
Sun
a
The disk shape of the star system reminded astronomers of the thick millstone disks used to grind flour.
b

▲**Figure 15-2** (a) In 1785, William Herschel published this diagram showing the shape of the star system as if it could be viewed edge-on from outside. (b) The Sun is located near the center of this "grindstone" model universe.



◀**Figure 15-3** (a) Open clusters lie along the Milky Way all around the sky. (b) Globular clusters are scattered over the entire sky but are strongly concentrated toward the direction of the galactic center. (c) Nearly half of cataloged globular clusters (*red dots*) are located in or near Sagittarius and Scorpius. A few of the brighter globular clusters, labeled with their catalog designations, are visible in binoculars or small telescopes. Constellations are shown as they appear above the southern horizon on a summer night as seen from latitude 40° N, typical for most of the United States.

hypothesized that the center of the star system could not be near the Sun but must lie somewhere toward Sagittarius.

To check his hypothesis, Shapley needed to find the distance to the center of the star system and thus the size of the star system as a whole. To do that, he needed to find the distances to individual star clusters, but that was difficult to do. The clusters are much too far away to have measurable parallaxes. They do, however, contain Cepheid variable stars (Sections 12-4a and b), which were the beacons Shapley needed to find the distances to the clusters.

Shapley knew about the prior work of Henrietta Leavitt on Cepheid variable stars. In 1912, Leavitt was studying a cloud of stars in the southern sky known as the Small Magellanic Cloud. On her photographic plates, she found many variable stars, and she noticed that the brightest had the longest periods. She didn't know the distance to the cloud, so she couldn't calculate absolute magnitudes (Section 9-2c), but because all of the variables she was observing are in the same star cloud and can be assumed to be at about the same distance, she concluded that there is a relationship between period and luminosity.

Based on Leavitt's work, Shapley and other astronomers realized that the Cepheids could be used to determine distances if their true luminosities could be discovered, but because Cepheids are giant and supergiant stars, they are relatively rare. None lies close enough to Earth to have a measurable parallax, but their proper motions—apparent slow movements across the sky because of their motions through space—can be measured (Section 9-1d). The more distant a star is, the smaller its proper motion tends to be, so proper motions contain clues to distance. Shapley found 11 Cepheids with measured proper motions, and then he used a statistical calculation technique to find their average distance and thus their average absolute magnitude. Then, he could erase Leavitt's *apparent* magnitudes from the vertical axis of the period–luminosity diagram (look back to Figure 12-14) and write in *absolute* magnitudes, representing true luminosities. All of the stars in the diagram were thus **calibrated** as standard light sources that astronomers could use to find distances. Calibrations like this one are important tools in science (**How Do We Know? 15-1**).

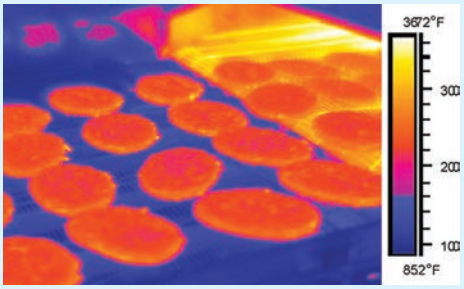
Calibration

How do you take the temperature of a vat of molten steel?

Astronomers often say that Shapley "calibrated" the Cepheids for the determination of distance, meaning that he did all the detailed background work to determine the luminosities of Cepheids. After that, he and other astronomers could use his calibrated period–luminosity diagram to find the distance to other Cepheids without repeating the calibration steps.

Calibration is common in science because it saves a lot of time and effort. For example, engineers in steel mills must monitor the temperature of molten steel, but they can't dip in a thermometer. Instead, they can use handheld devices that measure the color of molten steel. You may recall from Figure 7-7 that the color of blackbody radiation is determined by the temperature of the emitting object. Molten steel emits visible and infrared radiation that has a nearly perfect blackbody spectrum, so the manufacturer can calibrate the engineer's devices to convert the measured color to a temperature displayed on digital readouts. The engineers don't have to repeat the calibration every time; they just point their instrument at the molten steel and read off the temperature. Astronomers have made the same kind of color–temperature calibration for stars.

As you read about any science, notice how calibrations are used to simplify common measurements. But notice, too, how important it is to get the calibration right. An error in calibration can throw off every measurement made using that calibration.



Courtesy of FLIR Systems, Inc.

An infrared video camera calibrated to measure temperature allows bakers to monitor the operation of their ovens.

Finding distances using Cepheid variable stars is so important in astronomy that you should pause to examine the process. ***Example:*** Suppose you studied an open star cluster and discovered that it contained a type I Cepheid variable star with a period of 10 days and an average apparent magnitude of 10.5. How far away is the cluster? ***Solution:*** Using the period–luminosity diagram (Figure 12-14), you can see that the absolute magnitude of that star must be about −3.0. Now you can solve the magnitude–distance formula (Eq. 9-2b) to find the distance:

$$d = 10^{(m_V - M_V + 5)/5}$$

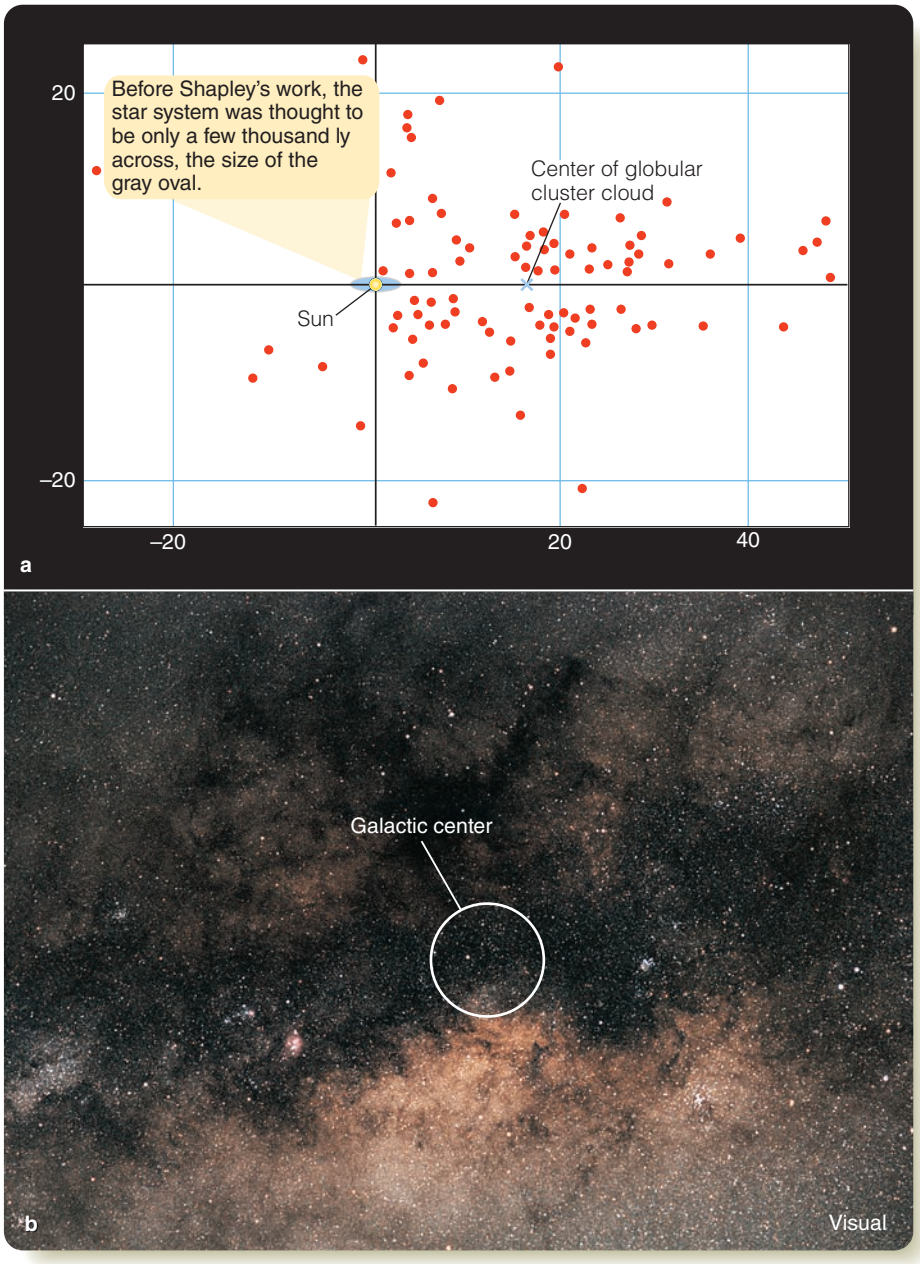
The difference between the apparent and absolute visual magnitude ($m_V - M_V$) is 10.5 minus (−3.0), which equals −13.5. If you use that value in the formula, you will get a result of $10^{3.7}$, or about 5000 pc. This is an example of one of the most common calculations in astronomy.

Once Shapley had calibrated the period–luminosity relation, he could use it to find the distance to any cluster in which he could identify those types of variable stars. By taking a series of photographic plates over a number of nights, Shapley was able to pick out the variable stars in a cluster, measure their average apparent magnitude, and find their periods of pulsation. Knowing that, he could read their absolute magnitude off the period–luminosity diagram. With the apparent magnitude and the absolute magnitude, he could calculate the distance to the star cluster.

This worked well for the nearer globular clusters, but variable stars in more distant clusters were too faint for him to detect. Shapley estimated the distances to these more distant clusters by calibrating the diameters of the clusters. For the clusters whose distance he knew, he could use their angular diameters and the small-angle formula (Eq. 3-1) to calculate linear diameters in parsecs. He found that the nearby clusters are about 25 pc in diameter, which he assumed is the average diameter of all globular clusters. He then used the angular diameters of the more distant clusters to find their distances. This is another illustration of calibration in astronomy.

Shapley later wrote that it was late at night when he finally plotted the directions and distances to the globular clusters on graph paper and found that, just as he had supposed, they formed a great swarm whose center lay many thousands of light-years away in the direction of Sagittarius, which confirmed his suspicion that the center of the star system was not near the Sun but was far away in Sagittarius (**Figure 15-4**). He found the only other person in the building, a cleaning lady, and the two stood looking at his graph as he explained that they were the only two people on Earth who understood that humanity lives, not near the center of a small star system, but in the suburbs of a vast wheel of stars.

It is important to note that there were problems with Shapley's calibrations and distance determinations. Shapley was able to observe the globular clusters at great distances because they lie outside the plane of the Galaxy (Figure 15-4a) and are not dimmed much by the interstellar medium, but the Cepheid stars he used for calibration mostly lie in the plane of the Galaxy and were strongly affected by extinction. Also, Shapley did not know there are several types of variable stars and that the ones he used for the distance estimates were different from the ones he used for the calibration. Consequently, his estimate for the size of the Galaxy was bigger than the modern value. Nevertheless, he got the main point right: We are thousands of light-years from the center of the Milky Way Galaxy.



Daniel Good

◀ **Figure 15-4** (a) Shapley's study of globular clusters showed that they were not centered on the Sun, at the origin of this graph, but rather form a great cloud centered far away in the direction of Sagittarius. Distances on this graph are given in thousands of parsecs, corresponding to Shapley's original calibration that produced distances more than two times larger than the modern values. (b) Looking toward Sagittarius at visual wavelengths, you see nothing to suggest that this is the center of our Galaxy. Interstellar dust and gas block your view. Only the distribution of globular clusters told Shapley that the center lies in this direction.

Building on Shapley's work, other astronomers began to suspect that some of the faint patches of light visible through telescopes were other star systems. Within a few years, they found evidence that the faint patches of light were indeed other galaxies much like our own Milky Way Galaxy. Today the largest telescopes can detect an estimated 100 billion galaxies similar to our own. Our home Galaxy is special only in that it is our home.

Shapley's study of star clusters led to the discovery that we live in a galaxy and that the Universe is filled with similar galaxies. Like Copernicus, Shapley moved us from the center to the outskirts. A careful analysis of the structure of the Galaxy will tell you the details about our real location.

15-2 Structure of the Galaxy

Astronomers commonly give the diameter of the Milky Way Galaxy as approximately 25 kpc, or 80,000 ly, and place the Sun about two-thirds of the way from the center to the edge (**Figure 15-5**). Note that this is only the diameter of the most visible part of our Galaxy. Later in this chapter you will learn that the entire Galaxy, some parts of which are not easy to detect, is much larger than 25 kpc in diameter.

15-2a Components of the Galaxy

The disk of the Galaxy is often referred to as the **disk component**. It contains most of the Galaxy's stars and nearly all

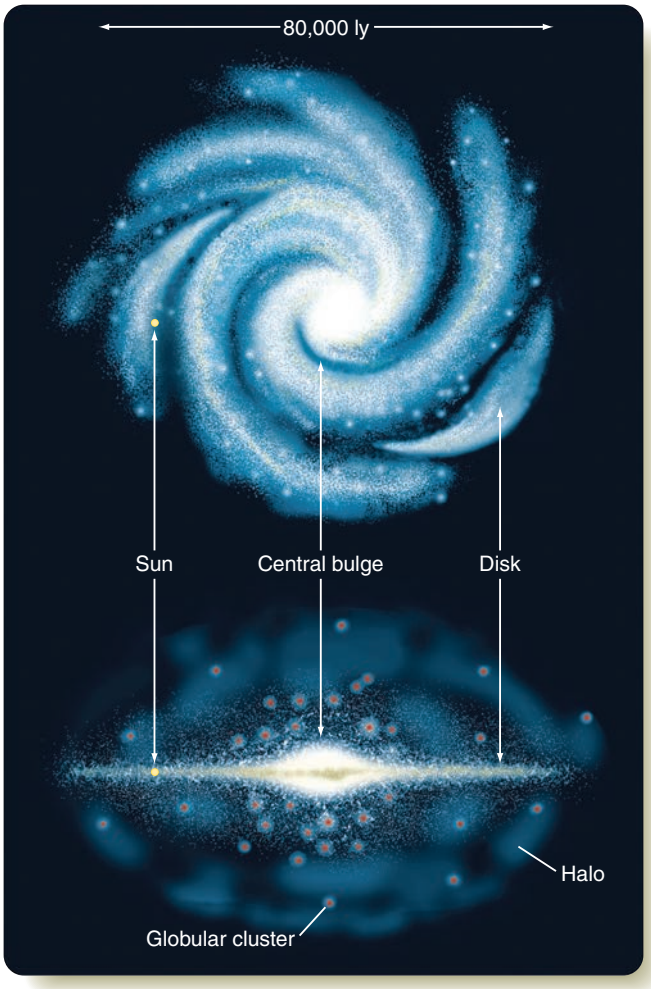
of its gas and dust. Because the disk is home to the giant molecular clouds within which stars form (look back to Section 11-1), nearly all of the star formation in our Galaxy takes place in the disk. Most of the stars in the disk are middle- to lower-main-sequence stars like the Sun, a few are red giants and white dwarfs, and a small number are brilliant blue O and B stars. Those hot, massive stars are rare, but they are so luminous that they provide much of the light from the disk.

Astronomers can't cite a single number for the thickness of the disk because it lacks sharp boundaries, and the thickness of the disk depends on the kind of object studied. Stars like the Sun, with ages of a few billion years, lie within a range of about 500 pc above and below the central plane. In comparison, the youngest stars, including the O and B stars, and the gas and dust from which stars form, are confined to a thin disk extending only about 50 pc above and below the plane (**Figure 15-6**). The 100 pc-thick layer of new stars and star-forming material has the same proportion to the 25,000 pc diameter of the Galaxy as the thickness of a piece of notebook paper compared to its width.

The disk of the Galaxy contains both stellar associations and open star clusters. Associations are groups of about ten to a few hundred stars so widely separated that their mutual gravity cannot hold them permanently together. From the turnoff points in their H–R diagrams (see the Concept Art: Star Clusters and Stellar Evolution in Chapter 12), you can tell that they are very young groups of stars. The stars move together through space (**Figure 15-7**) because they formed from a single interstellar cloud and haven't yet had time to wander apart since they formed.

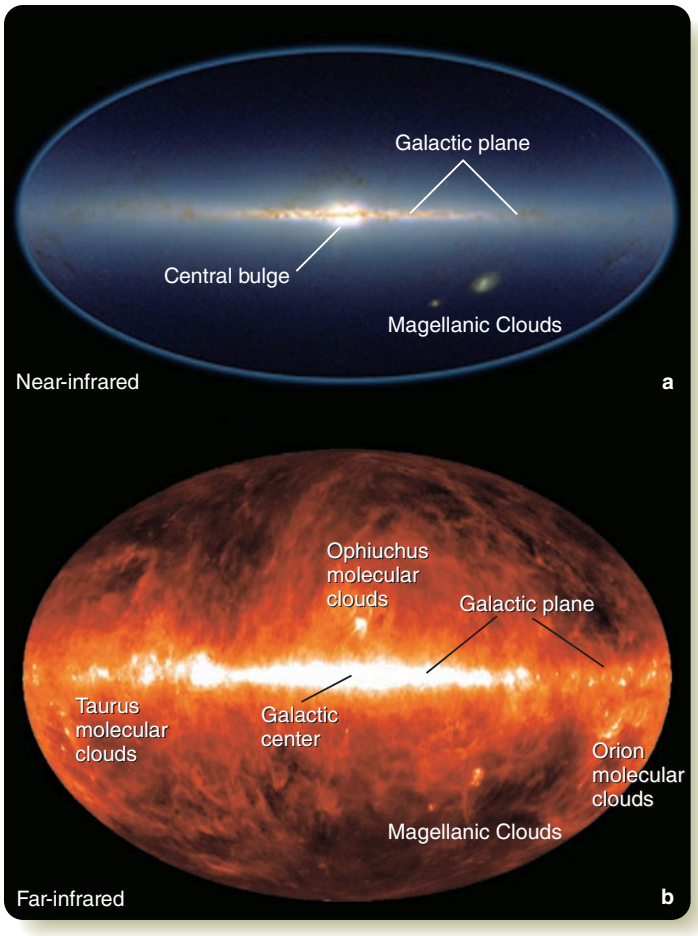
The second kind of star group found in the disk is the open cluster (Figure 15-3a), a group of 100 to a few thousand stars in a region about 25 pc in diameter. Because they have more stars crowded into smaller volumes than associations, open clusters are more firmly bound by gravity. Although they lose stars occasionally, they can survive for a long time, and the turnoff points in their H–R diagrams give ages from a few million to a few billion years.

If you look "up" or "down," out of the Galaxy's disk, you are looking away from the dust and gas, so you can see into the



© 2003, 2004 Three Rivers Foundation

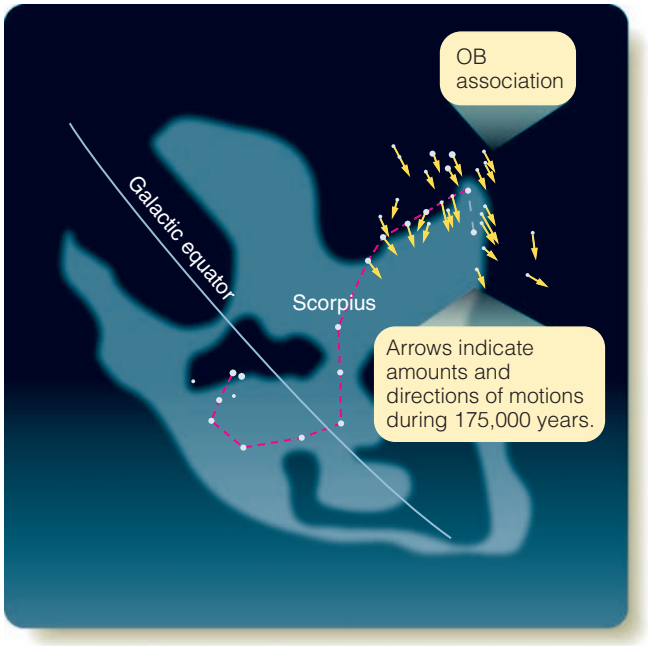
▲ **Figure 15-5** Sketches of our Milky Way Galaxy viewed face-on and edge-on. Note the position of the Sun and the distribution of globular clusters in the halo. Hot, blue stars light up the spiral arms. Only the inner halo is shown here. At this scale, the entire halo would be larger than a dinner plate.



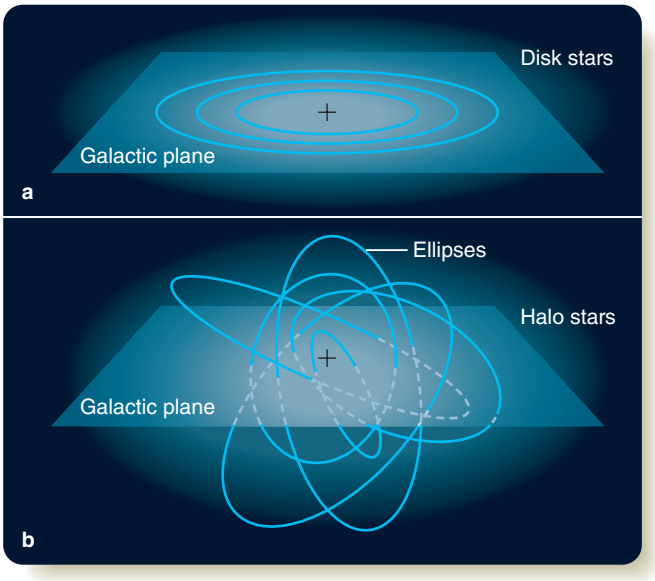
NASA/NSF/2MASS/IPL-Caltech/Univ. of Massachusetts

NASA/*COBE*/DIRBE, courtesy of H. Freudenreich

▲ **Figure 15-6** In these infrared images, the entire sky has been projected onto ovals with the center of the Galaxy at their respective centers. The central plane of the Milky Way extends from left to right. (a) The central bulge is prominent at near-infrared wavelengths, but dust clouds block the view through the plane. (b) Interstellar dust emits blackbody radiation and glows brightly at longer wavelengths.



▲ **Figure 15-7** Many of the stars in the constellation Scorpius are members of an OB association that evidently formed recently from a single cloud of gas. As the stars orbit the center of our Galaxy, they are moving together southwest along the Milky Way.



▲ **Figure 15-8** (a) Stars in the galactic disk, including the Sun, have nearly circular orbits that lie in the plane of our Galaxy. (b) Stars in the halo have randomly oriented, eccentric orbits.

Galaxy's **halo**, which is a spherical cloud of stars and star clusters that contains almost no gas and dust. Because the halo contains no dense gas clouds, it cannot make new stars. Halo stars are old, cool, lower-main-sequence stars; red giants; and white dwarfs. It is difficult to judge the extent of the halo, but it is estimated to have as much as ten times the diameter of the visible disk.

At the center of our Galaxy lies the **central bulge**, which is a flattened cloud of billions of stars about 6 kpc in diameter (Figure 15-5) that is visible above and below the plane of the Milky Way and through gaps in obscuring interstellar dust (Figure 15-4b). Like the halo, the central bulge itself contains little gas and dust. Astronomers often refer to the halo and the central bulge together as the **spherical component** of the Galaxy. The central bulge is the most crowded part of the spherical component, consisting mostly of stars that are old and cool like the stars in the halo.

The halo includes more than 150 globular clusters, each of which contains 50,000 to a million stars within a sphere a few tens of parsecs in diameter (Figure 15-3b). Astronomers can tell their average age is about 11 billion years from the positions of the turnoff points in their H–R diagrams.

Orbital motions are quite different in the two components of the Galaxy. Disk stars follow nearly circular orbits that lie in the plane of the Galaxy (**Figure 15-8a**). Halo stars and globular clusters, however, follow highly elongated orbits tipped steeply to the plane of the disk (**Figure 15-8b**). (Although the diagram shows these orbits as simple ellipses, that is not quite true. The gravitational influence of the thick, asymmetric bulge forces them into curves that do not quite return to the same starting point.) The dramatic difference between the motions of halo stars and disk stars will be important evidence when you consider the formation of the Galaxy later in this chapter.

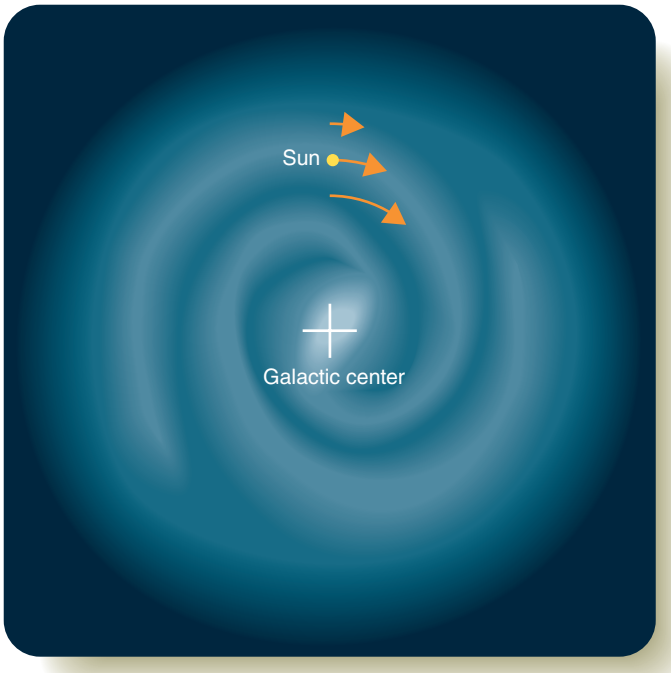
This analysis of the components of the Milky Way Galaxy leads to an important question: How much matter does our Galaxy contain? To answer that question, you can measure the Galaxy's rotation.

15-2b Mass of the Galaxy

To find the mass of an object, astronomers must observe the motion of another object orbiting it, just as in a binary star system. Humans don't live long enough to see stars move significantly along their orbits around the Galaxy, but astronomers can observe the radial velocities, proper motions, and distances of stars and then calculate the sizes and periods of their orbits. The results reveal the mass of the Galaxy.

Stars in the disk of the Galaxy follow nearly circular orbits that lie in the plane of the disk (look again at Figure 15-8a). Based on current observations, the Sun orbits the center of our Galaxy at about 225 km/s and moves in the direction of Cygnus. The evidence indicates that the Sun's orbit is nearly circular; so given the current best estimate for the distance to the center of our Galaxy, 8.3 kpc, you can find the circumference of the Sun's orbit by multiplying by 2π. If you divide the circumference of its orbit by its orbital velocity, you will discover that the Sun has an orbital period of about 225 million years.

When you studied binary stars in Section 9-5b, you learned that the total mass M (in solar masses, $M_\odot$) of a binary star system equals the cube of the average separation of the stars a (in AU) divided by the square of the period P (in years). (Recall that this is Kepler's third law for stellar masses, Eq. 9-4.) ***Example:*** What is the mass of the Galaxy? ***Solution:*** The radius of the Sun's orbit



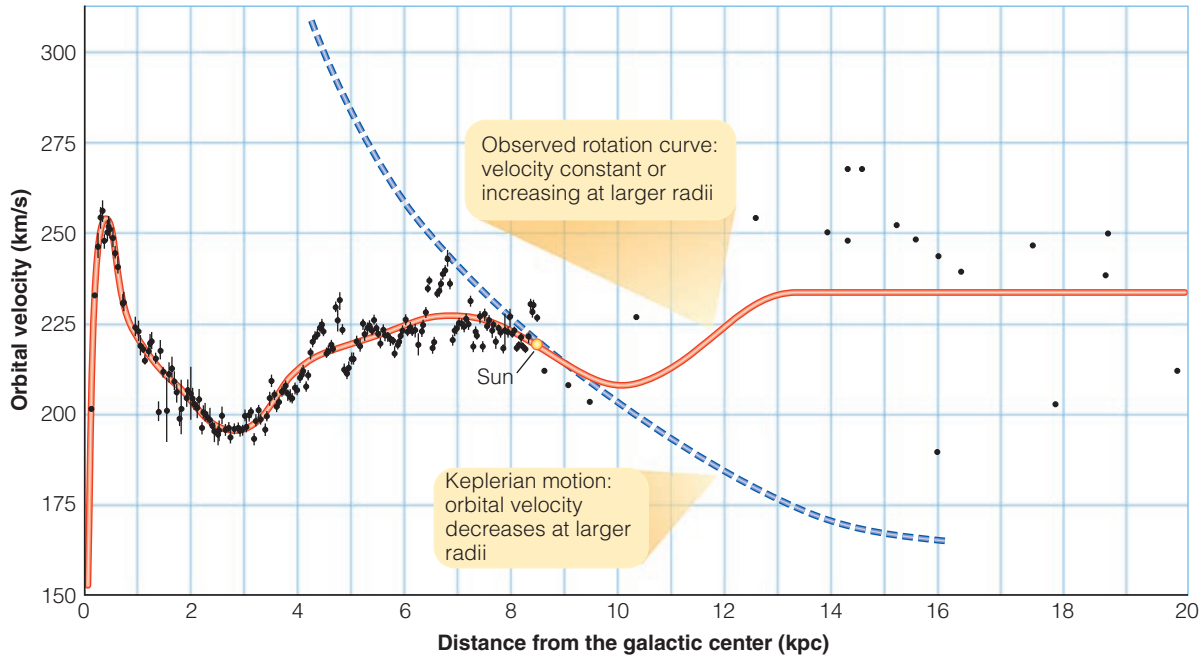
▲ **Figure 15-9** The differential rotation of our Galaxy means that stars at different distances from the center have different orbital periods. In this example, the star just inside the Sun's orbit has a shorter period and pulls ahead of the Sun, whereas the star outside falls behind.

around the center of the Milky Way Galaxy is about 8300 pc, and each parsec contains 2.1×10^5 AU (to a precision of two digits). Multiplying, you find that the radius of the Sun's orbit is 1.7×10^9 AU. The orbital period is 225 million years, so the complete calculation tells you that the mass of the Galaxy is 1.0×10^{11} (100 billion) $M_\odot$. This is a rough estimate because it only includes the mass lying within the Sun's orbit. Allowing for that overlooked mass yields a lower limit to the total mass for our Galaxy of at least 4×10^{11} (400 billion) $M_\odot$.

The rotation of our Galaxy is actually the orbital motion of each of its stars around the center of mass. Stars at different distances from the center revolve around the center of the Galaxy with different periods, so stars starting near each other will draw apart as time passes (**Figure 15-9**), which is called *differential rotation*. (Recall that a somewhat different meaning of differential rotation was used regarding the Sun in Section 8-2c.)

To fully describe the rotation of our Galaxy, astronomers graph orbital velocity versus radius to produce a **rotation curve** (**Figure 15-10**). If all of the mass of the Galaxy were concentrated near its center, you would expect to see orbital velocities decrease for objects farther away from the center. That is what you see in our Solar System, where nearly all of the mass is concentrated in the Sun. For that reason, this is called **Keplerian motion**, referring to Kepler's laws. In contrast, the best observations of the rotation curve of the Milky Way show that orbital velocities in the outer disk are constant or even increasing at greater distances from the center of the Galaxy. Those high orbital velocities indicate that the larger orbits enclose more mass and imply that our Galaxy has much more mass than is contained within the radius of the Sun's orbit.

Both observational evidence and mathematical models show that the extra mass lies in an extended halo, sometimes called a



▲ **Figure 15-10** The rotation curve of our Galaxy is plotted here as orbital velocity versus radius. Data points show measurements made by radio telescopes. Observations outside the orbit of the Sun are much more uncertain, and the data points scatter widely. Orbital velocities do not decline outside the orbit of the Sun, as you would expect if most of the mass of our Galaxy were concentrated toward the center (Keplerian motion). Rather, the curve is approximately flat at great distances, suggesting that our Galaxy contains significant invisible mass ("dark matter") outside the orbit of the Sun.

galactic corona, that may reach out to ten times farther than the edge of the visible disk and could contain more than a trillion (10^{12}) $M_\odot$. Most of that mass is invisible, neither emitting nor absorbing light, so astronomers refer to it as **dark matter**. Some of the mass in the galactic corona is made up of low-luminosity stars and white dwarfs, but much of the mass must be some other form of matter. You will learn more about the problem of dark matter in the following two chapters. The nature of dark matter is one of the most important unknowns in modern astronomy.

Practicing SCIENCE

If dust and gas block the view, how do astronomers know how big our Milky Way Galaxy is? This is just one example of the type of question scientists must continually ask themselves as they practice their profession: How do we know what we know?

Interstellar dust in our Galaxy blocks the view only in the plane of the Galaxy. When you use a telescope to look away from the plane of the Galaxy, you are looking out of the obscuring interstellar medium and can see to great distances. The globular clusters are scattered through the halo, with a strong concentration in the direction of Sagittarius. When astronomers look above or below the plane of the Galaxy, they can see those clusters, and careful observations reveal variable stars in the clusters. They can then find the distances to those clusters by using modern calibrations of the period–luminosity diagrams of Cepheid and RR Lyrae variable stars found in those clusters. Astronomers infer that the distance to the center of the distribution of globular clusters is the distance to the center of the Galaxy because they assume that the distribution of clusters is controlled by the gravitation of the Galaxy as a whole.

Regarding any fact, it is important to ask "How do we know?" Now review how we know another important fact about the Milky Way: ***What measurements and assumptions are needed to determine the total mass of our Galaxy?***

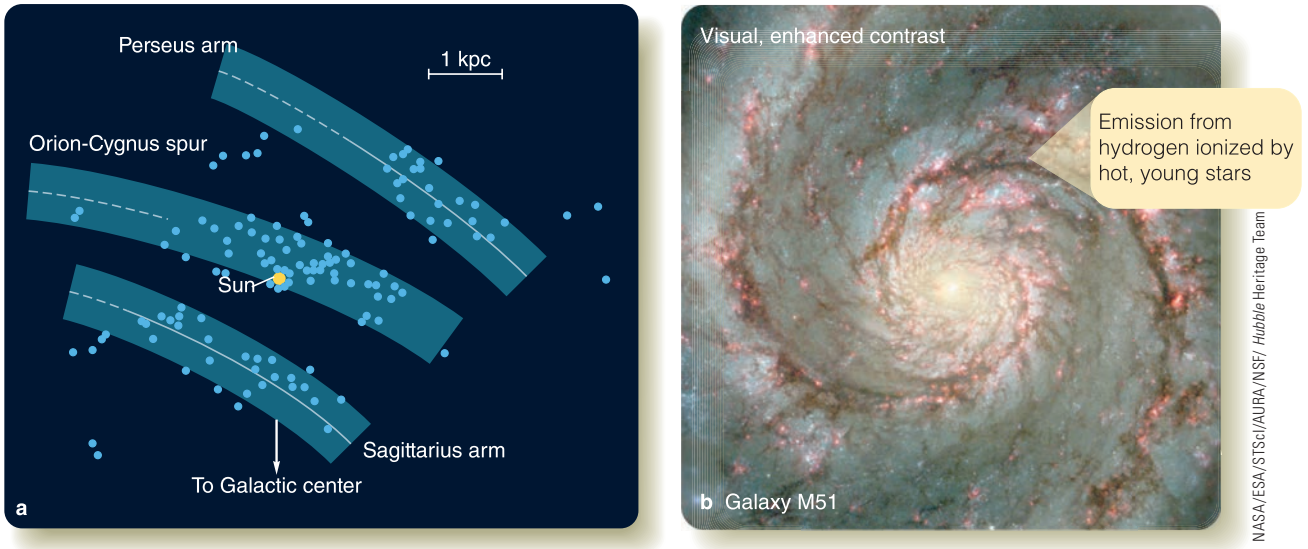
15-3 Spiral Arms and Star Formation

The most striking feature of galaxies like the Milky Way is their patterns of **spiral arms** that wind outward through the disk. These arms contain swarms of hot, blue stars; clouds of dust and gas; and young star clusters. The young objects suggest that the spiral arms involve star formation, but as you try to understand the spiral arms, you need to consider two problems. First, how can anyone be sure our Galaxy has spiral arms if interstellar dust obstructs our view? Second, why doesn't the differential rotation of the Galaxy destroy the arms? The solution to both problems involves star formation.

15-3a Tracing the Spiral Arms

As you have already learned, O and B stars are often found in associations and are very luminous. Therefore, they can be detected across great distances. At those great distances their parallaxes and other properties are difficult to measure, but the available data, especially from infrared and radio observations, indicate that OB associations in the region of the Milky Way near the Sun are not located randomly but lie along several arcs that are understood to be segments of spiral arms (**Figure 15-11a**). (The dust blocking our view at visual wavelengths is transparent to infrared and radio waves because those wavelengths are much larger than the diameter of the dust particles.) Those spiral segments have been named for the prominent constellations through which they pass.

Objects used to map spiral arms are called **spiral tracers**. Aside from OB associations, spiral tracers include young open star clusters, emission nebulae (clouds of hydrogen ionized by hot stars), and certain high-mass variable stars. Notice that all of these spiral tracers are young objects, formed recently, astronomically speaking. O stars, for example, live for only a few million years. Their typical galactic orbital velocities, like the Sun's, are about 200 km/s, so they cannot move more than about 500 pc in their lifetimes. This is less than the



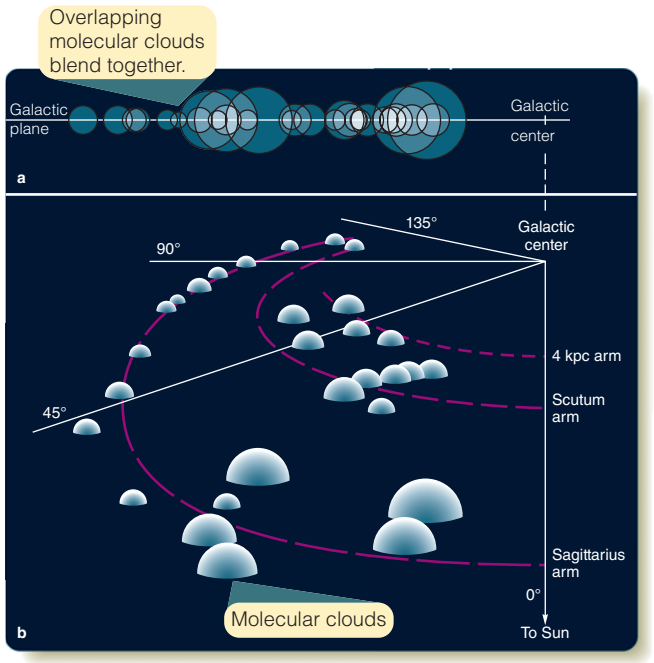
▲ **Figure 15-11** (a) Nearby young O and B stars in our Galaxy fall along bands that appear to be segments of spiral arms. (b) Images of other galaxies show that spiral arms are marked especially by hot, luminous stars and emission nebulae that must be very young. This strongly suggests that spiral arms are related to star formation.

width of a spiral arm. Because they don't live long enough to move away from the spiral arms, they must have formed there.

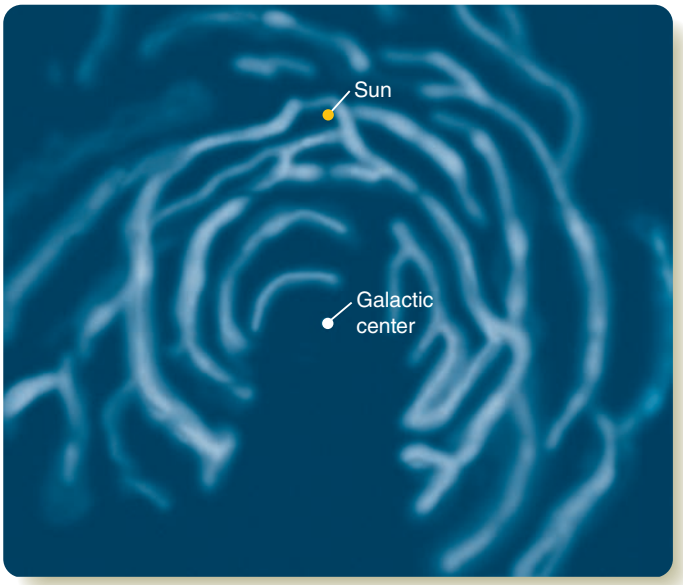
Studies of other galaxies show that their spiral arms are also marked by hot blue stars and other tracers like the ones in the Milky Way Galaxy (**Figure 15-11b**). The youth of spiral tracers is an important clue about spiral arms. Obviously spiral arms are associated with star formation. But, before you can follow this clue, you need to extend your map of spiral arms to show the entire Galaxy.

15-3b Radio Maps of Spiral Arms

Radio astronomers often use the strong spectral line emission from carbon monoxide (CO) to map the location of giant molecular clouds in the plane of the Galaxy. Recall from Section 11-1a that giant molecular clouds are sites of active star formation, so you can expect them to be connected with the other spiral tracers. If you point a radio telescope at a section of the Milky Way, you will receive a combination of signals from gas clouds that lie at various distances across the Galaxy in the direction you are looking. Signals from the different clouds can be distinguished by measuring the Doppler shifts of their spectral lines. Astronomers then use a simple model of orbital velocities at different distances from the center of the Galaxy to untangle the observations and locate the individual clouds. Maps constructed from such observations reveal that, in fact. the giant molecular clouds, like O and B stars, are located along segments of spiral arms (**Figure 15-12**).



▲ **Figure 15-12** (a) At the wavelengths of spectral lines emitted by carbon monoxide molecules, astronomers find many molecular clouds along the Milky Way, but the clouds overlap in a confusing jumble. By using a model of the rotation of our Galaxy to interpret the radial velocities of the clouds, radio astronomers can estimate the distances to each cloud and use them to map spiral arms. (b) This map represents the view from a point 2 kpc directly "above" the Sun (toward galactic North). The molecular clouds, shown here as hemispheres extending above the plane of our Galaxy, are located along spiral arms. Angle labels indicate galactic longitudes.



▲ **Figure 15-13** This 21-cm radio map of our Galaxy confirms that concentrations of neutral atomic hydrogen gas lie approximately in a spiral, but the pattern is complex and suggests branches and spurs.

You learned in Section 10-1f that observing at a wavelength of 21 cm allows detection of a spectral line of atomic hydrogen. That type of gas is generally warmer than the material in molecular clouds. Just as for the molecular clouds, analysis of the atomic cloud radial velocity data requires that astronomers start with estimates of the orbital velocities at different distances from the center of the Galaxy. You can see that the map of warm atomic hydrogen gas (**Figure 15-13**) doesn't trace as clear a spiral pattern as the cold molecular CO gas does. That is because astronomers have difficulty determining radial velocities of the atomic gas clouds, which have more random motions caused by higher temperatures and greater turbulence than in molecular clouds. Radial velocity measurements toward the center of the Galaxy have an additional problem. There, orbital motions of gas clouds are perpendicular to the line of sight, and all of the radial velocities are zero. That is why the map in Figure 15-13 is empty in the wedge-shaped region directly toward the center.

Clearly we live in a spiral galaxy, but the spiral pattern appears to be slightly irregular, with many branches and gaps. The luminous and star-forming nebulae you see in Orion, for example, may be part of a detached segment of a spiral arm, referred to as a "spur." By carefully combining observations of our own Galaxy with studies of other galaxies, astronomers can infer what the Milky Way Galaxy might look like viewed from outside, as in the artist's conception that opens this chapter. In addition to helping map the arrangement of spiral arms, infrared observations indicate that the central bulge of our Galaxy is not a sphere but rather an elongated bar pointing partially away from Earth. In the next chapter, you will see that such structures are common in other galaxies.

One important fact revealed by the radio maps is that spiral arms are regions of higher gas density. Spiral tracers are all comparatively young objects, so spiral arms are obviously sites of

active star formation. Radio maps are consistent with that, showing that the material needed to make stars is abundant and concentrated in spiral arms.

15-3c The Spiral Density Wave Theory

Having mapped the spiral pattern of our Galaxy and seen them in other galaxies, you might ask, "Just what are spiral arms?" You can be sure they are not physically connected structures such as bands of gas and stars held together by their own gravity. If they were, the differential rotation of a galaxy would destroy them within an astronomically short time, just a few hundred million years, winding up the bands and tearing them apart like paper streamers caught on a rotating wheel. Because spiral arms are common in galaxies, they must last for billions of years.

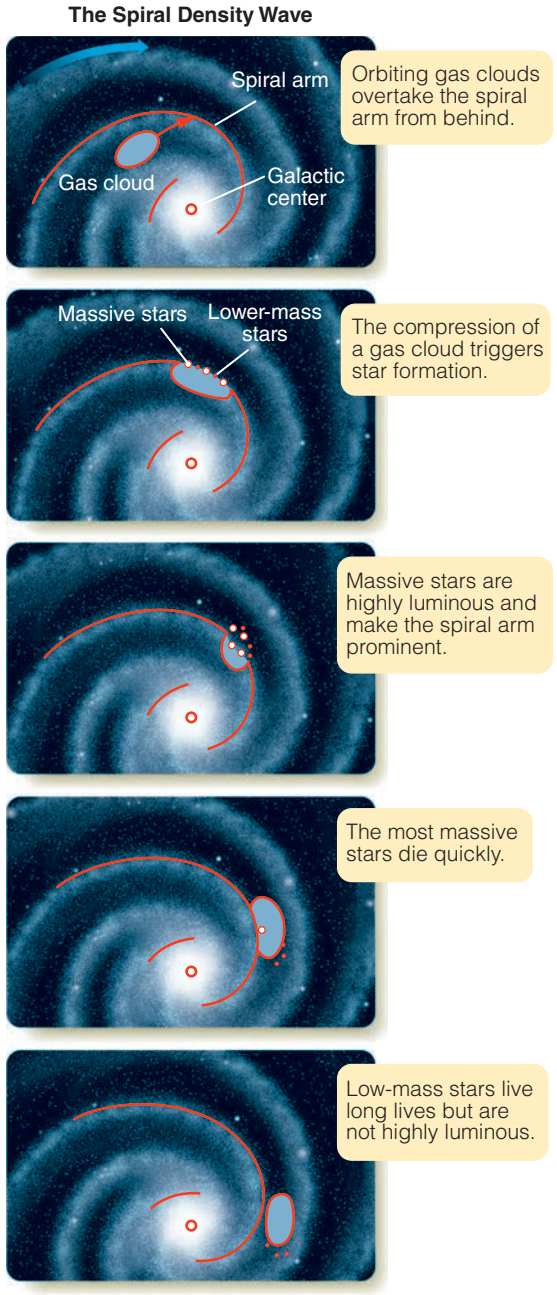
Astronomers conclude that spiral arms are dynamically stable—they retain the same appearance even though the gas, dust, and stars in them are constantly changing. To see how this could work, think of the traffic jam behind a slow-moving truck. Seen from a traffic helicopter, the jam would be stable, moving slowly down the highway. But you could watch an individual car approach from behind, slow down, wait its turn, finally reach the front of the jam, pass the truck, and resume speed. The individual cars in the jam are constantly changing, but the traffic jam itself is a stable pattern, moving at its own speed along the highway.

In the **spiral density wave theory**, spiral arms are dynamically stable regions of compressed interstellar medium that move slowly around the Galaxy, just as the truck moves slowly down the highway. Gas clouds moving at orbital velocity around the Galaxy overtake the slow-moving arms from behind and slam into the gas already in the arms. The resulting sudden compression can trigger the collapse of the gas clouds and the formation of new stars (Section 11-1). Newly formed stars and the remaining gas eventually move on through the arm and emerge from the front of the slow-moving arm, continuing their travels around the Galaxy (**Figure 15-14**).

This model explains the existence of spiral tracers. Stars of all masses are forming in spiral arms, but, as you have already learned, the O and B stars live such short lifetimes that they die before they can move out of the spiral arm. The presence of such massive, high-luminosity stars, along with dense molecular clouds, confirms that spiral arms are the primary sites of star formation.

Plotting the locations of Sun-like stars does not reveal any spiral pattern, but it's easy to understand why lower-mass stars like the Sun can't be used as spiral tracers. Such stars also form in spiral arms, but during their relatively long lives, they leave the arms in which they were born. The Sun formed 4.6 billion years ago as part of an association in a star-forming region within a spiral arm, escaped from that association, and has circled the Galaxy more than 20 times, passing through many spiral arms.

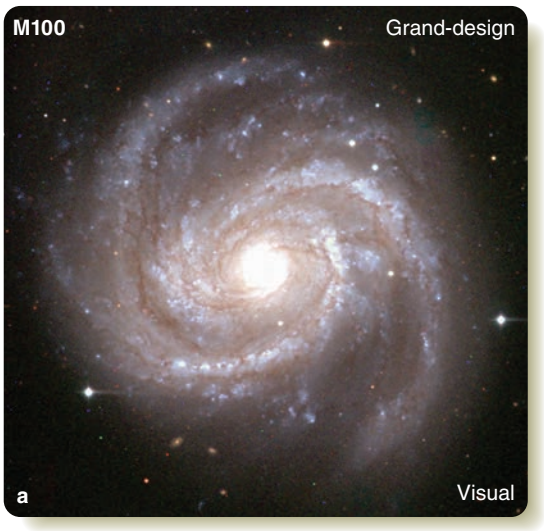
The evidence seems to fit the spiral density wave theory well, but the theory has two problems. First, how does the enormous and complicated spiral disturbance begin, and how is it sustained? Computer models indicate that spiral density waves should fade away in about a billion years, so something must

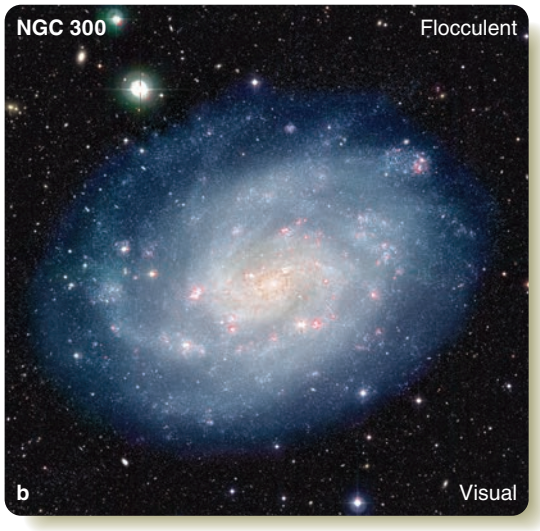

The Spiral Density Wave
Spiral arm
Orbiting gas clouds overtake the spiral arm from behind.
Gas cloud
Galactic center
Massive stars
Lower-mass stars
The compression of a gas cloud triggers star formation.
Massive stars are highly luminous and make the spiral arm prominent.
The most massive stars die quickly.
Low-mass stars live long lives but are not highly luminous.

▲ **Figure 15-14** Artist's conception: According to the spiral density wave theory, star formation occurs as gas clouds pass through spiral arms.

regenerate the spiral wave. Other models show that the disk of the Galaxy is strongly affected by certain types of disturbances. As you learned previously, observations show that the center of our Galaxy is not a sphere but a bar. The gravitational effects of the bar's rotation could continuously perturb the Galaxy's disk and stimulate the formation of spiral density waves. A close encounter with a passing galaxy could also stimulate a spiral pattern. However, some galaxies without a visible bar or large companion nevertheless have prominent spiral arms, so astronomers continue to make observations and computer models, investigating questions of how spiral arms form and are sustained.

The second problem for the density wave theory involves the spurs and branches observed in the arms of our own and other





ING/J. E. Beckman, R. F. Peletier, J. H. Knapen, R. L. M. Corradi, L. J. Gentet

ESO/M. Schirmer (IAEF, Bonn), W. Gieren (Univ. de Concepción, Chile), et al.

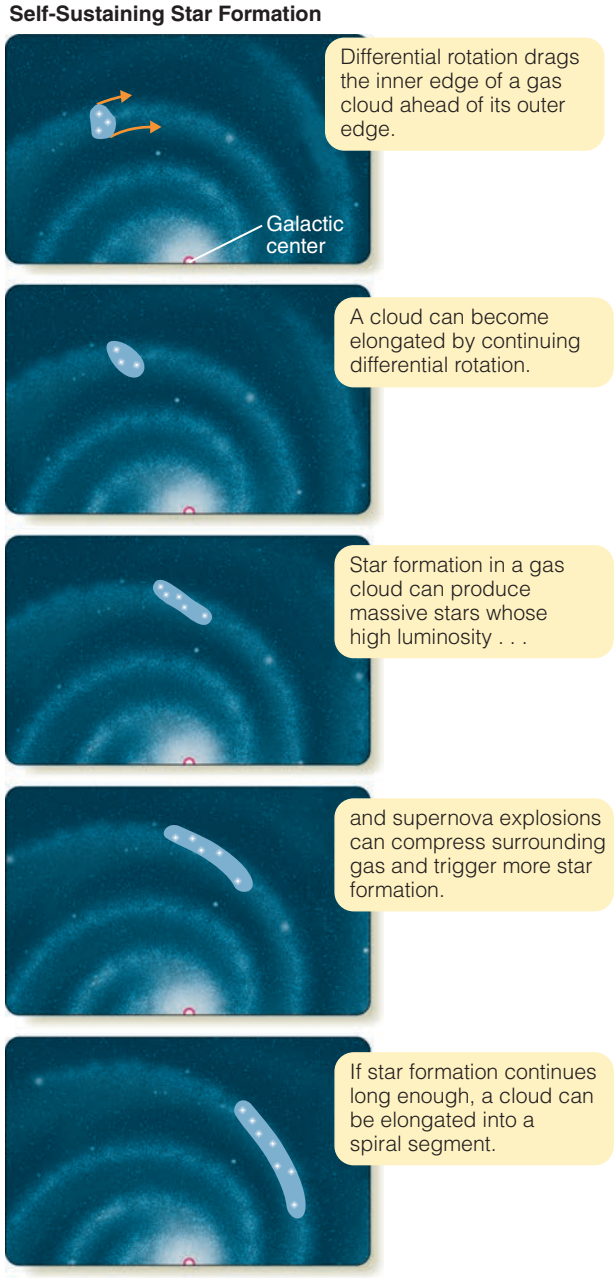
▲ **Figure 15-15** Images of spiral galaxies show that spiral arms are marked especially by hot, luminous stars and emission nebulae related to star formation. (a) Some galaxies are dominated by two spiral arms, but, even in these galaxies, minor spurs and branches are common. Spiral density waves can generate the two-armed, grand-design pattern, but self-sustained star formation may be responsible for the irregularities. (b) Many spiral galaxies have disconnected spiral segments, spurs, and branches but no arms that extend across the galaxy. Nevertheless, star formation is evidently proceeding robustly in such galaxies, termed flocculent spirals. Observations indicate that our Milky Way Galaxy's spiral pattern is intermediate between these two examples.

galaxies. Some galaxies, called **grand-design galaxies**, have bold, symmetric two-armed patterns (Figure 15-15a). Other galaxies have a great many short spiral segments, giving them a fluffy appearance, but no overall grand design. These galaxies have been termed **flocculent galaxies**, meaning "woolly" (Figure 15-15b). Our Milky Way Galaxy seems to be intermediate between the extremes of flocculent and grand design.

15-3d Self-Sustaining Star Formation

What process might make small-scale spiral segments in a flocculent galaxy instead of a galaxy-wide spiral pattern? The answer may lie in self-sustaining ("contagious") star formation. You learned in Section 11-2b that star formation in one location can cause star formation in neighboring locations via supernova explosions, stellar winds, and outflows (Figures 11-5 and 11-6). The differential rotation of a galaxy will drag the inner edge of a star-forming cloud (closest to that galaxy's center) ahead, and let the outer edge lag behind. This will result in a region of star formation shaped like a segment of a spiral arm: a spur (Figure 15-16). Astronomers suspect that although a spiral density wave can generate beautiful two-armed patterns, the **self-sustaining star formation** process produces the branches and spurs that are prominent in some galaxies, including our own.

This discussion of star formation in spiral arms illustrates the importance of natural processes. The spiral density wave creates the


Self-Sustaining Star Formation
Differential rotation drags the inner edge of a gas cloud ahead of its outer edge.
Galactic center
A cloud can become elongated by continuing differential rotation.
Star formation in a gas cloud can produce massive stars whose high luminosity . . .
and supernova explosions can compress surrounding gas and trigger more star formation.
If star formation continues long enough, a cloud can be elongated into a spiral segment.

▶ **Figure 15-16** Self-sustaining star formation may be able to produce long clouds of young stars that look like segments of spiral arms.

Nature as Processes

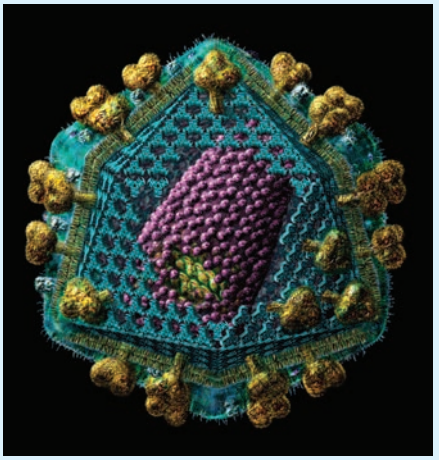
How is getting a cold like stars building the chemical elements?

Science, at first glance, seems to be nothing but facts, but in many cases, you can organize the facts into the story of a process. For example, astronomers try to assemble the sequence of events that led to the formation of the chemical elements. If you understand that process, you have command over a lot of important facts in astronomy.

A process is a sequence of events that leads to some result or condition, and much of science is focused on understanding how these natural processes work. Biologists, for example, try to understand how a virus reproduces. They must figure out how the virus tricks the immune system into ignoring it, penetrates the wall of a healthy cell, injects viral DNA, commandeers the cell's resources to make new viruses, and finally destroys the cell to release the new virus copies. One biologist may spend years studying a specific step, but the ultimate goal is to be able to tell the entire story of the process.

As you study any science, be alert for processes as organizing themes. When you see a process in science, ask yourself a few basic questions. What conditions prevailed at the beginning of the process? What sequence of steps occurred? Can some steps occur simultaneously, or must one step occur before another can occur? What is the final state that this process produces?

Recognizing a process and learning to tell its story will help you remember a lot of details, but there is more to understanding a process than its value as a memory aid. Identifying a process and learning to tell its story helps you understand how nature works and why the Universe is the way it is.



Russell Knightly Media, rkm.com.au

A virus is a collection of molecules that cannot reproduce until it penetrates into a living cell.

arms, but it is the star formation in the arms that makes them stand out so prominently. Self-sustaining star formation can act in some galaxies to modify the spiral arms and produce branches and spurs. In other galaxies, it can make the spiral pattern flocculent. By searching out and understanding the details of such natural processes, astronomers can begin to understand the overall structure and evolution of the Universe we live in (**How Do We Know? 15-2**).

Practicing **SCIENCE**

Why can't astronomers use Sun-like stars as spiral tracers?

One important part of practicing science is being able to choose between observations that will address a certain question from observations that will be less helpful or even increase confusion.

In this example, you need to consider the evolution of stars and their orbital periods around the Galaxy. Stars like the Sun live about 10 billion years, but the Sun's orbital period around the Galaxy is 225 million years. The Sun almost certainly formed when a gas cloud passed through a spiral arm, but since then the Sun has circled our Galaxy many times and has passed through spiral arms often. That means the Sun's present location is only coincidentally near a spiral arm. An O star, however, lives only a few million years. It is born in a spiral arm and lives out its entire lifetime before it can leave the spiral arm. Short-lived stars such as O stars are found only in spiral arms, but G stars are found all around our Galaxy. Careful measurements of the positions of O stars can reveal the spiral structure of our Galaxy, but careful measurements of the positions of G stars will not help at all.

The spiral arms of our Galaxy would make it beautiful if it could be photographed from a distance, but we are trapped inside it. Expand the question about observing the spiral structure of our Galaxy: ***What measurements are needed to show that the spiral arms mapped by nearby spiral tracers actually extend across the disk of the Milky Way Galaxy?***

15-4 The Nucleus of the Galaxy

The most mysterious region of our Galaxy is its very center, the nucleus. At visual wavelengths, this region is totally hidden by interstellar dust that dims the light from there by 30 magnitudes. If a trillion (10^{12}) photons of light left the center of the Galaxy on a journey to Earth, only one of those photons would make it through the dust. Consequently, visual-wavelength images reveal nothing about the nucleus (Figure 15-4b). Observations at infrared and radio wavelengths can see through the interstellar material, and those images show a region of tremendously crowded stars orbiting in the nucleus at high velocities along with complicated structures of atomic and molecular gas clouds. To understand what is happening in the innermost regions of our Galaxy, you need to carefully compare observations and hypotheses.

15-4a Observations of the Nucleus

If you look up at the Milky Way on a dark night, you might notice a slight thickening in the direction of the constellation Sagittarius, but nothing specifically identifies this as the direction of the heart of the Galaxy. Even Shapley's study of globular clusters identified the location only approximately. The first complete infrared map of the central bulge made by Eric Becklin in 1968 showed the location of the strongest glow where the stars are most crowded together—the gravitational core of the Galaxy. Later, high-resolution radio maps of that stellar center revealed an abundance of radio sources, with one, Sagittarius A* (Sgr A*)—usually pronounced *sadge A-star*—lying at the very center of the galactic nucleus.

NASA/CXC/SAO; Visual: NASA/ESA/STScI/AURA/NSF; Infrared: NASA/SSC/JPL-Caltech

◀ **Figure 15-17** Data from the *Hubble Space Telescope,* the *Spitzer Space Telescope,* and the *Chandra X-ray Observatory* were combined to produce this image of the central 70 pc (230 ly) of our Milky Way Galaxy. This region evidently contains a turbulent interstellar medium and powerful energy sources. The exact center of the Galaxy is located within the bright white region at lower right. If it could be seen from Earth with the unaided eye, the area shown in this image would be about half the angular diameter of the Moon.

Radio interferometer observations show that Sgr A* is less than 1 AU in diameter. Nevertheless, it is a strong source of both radio energy and X-ray emission (short-wavelength X-rays can also penetrate the interstellar medium). **Figure 15-17** is an image of the central 70 pc of the Galaxy that combines data from three space telescopes: near-infrared (represented by yellow), mid-infrared (red), and X-ray (blue and green). You can see that powerful forces are at work in the nuclear region. The tremendous amount of infrared radiation coming from the central area appears to be produced by crowded stars and by dust warmed by those stars. But what could be as small as Sgr A* and produce so much radio and X-ray energy?

Read **Sagittarius A*** and notice three important points:

1. Observations at radio and infrared wavelengths reveal complex structures near Sgr A* caused by magnetic fields and by vigorous star formation. Supernova remnants show that massive stars must have formed there recently and exploded at the ends of their lives.
2. The nucleus of the Galaxy is crowded. Tremendous numbers of stars heat interstellar dust, which emits strong infrared radiation.
3. There is evidence that Sgr A* is a supermassive black hole into which gas is flowing. Observations of the motions of stars orbiting the central object indicate its mass is at least 4 million $M_{\odot}$.

A supermassive black hole is an exciting idea, but scientists try always to be aware of the difference between possibility and likelihood. A supermassive black hole possibly can explain the observations, but is it likely? Could there be other explanations? For example, some astronomers suggested that gas flowing inward from the galactic halo could trigger tremendous bursts of star formation, causing some of the turmoil observed in the galactic nucleus. Others suggested that a dense cluster of millions of stellar-mass neutron stars and black holes could produce the same effects as one supermassive black hole. Such hypotheses have been considered and tested against the evidence, but none appears to be adequate to explain all the observations.

So far, the only hypothesis that explains all the observations is that our Galaxy's nucleus is home to a supermassive black hole. Astronomers are planning to study Sgr A* with an interferometer array (look back to Section 6-5d) consisting of many of the world's most powerful radio telescopes. Dubbed the Event Horizon Telescope, this array should be capable of resolving inflowing material right at the edge of the black hole.

Meanwhile, less ambitious observations are allowing astronomers to refine their models of the nucleus of the Milky Way. For instance, Sgr A* is not as bright in X-rays as it should be if it has a hot accretion disk with matter constantly flowing into the black hole. Right now Sgr A* seems to be mostly dormant, on a "starvation diet." Observations of X-ray and infrared flares lasting only a few hours suggest, however, that Sgr A* may occasionally go off its diet as mountain-size blobs of gas are ripped apart and heated by tides as they fall in the hole.

Such a supermassive black hole could not be the remains of a single dead star; it is far too large. In later chapters, you will see that such supermassive black holes are found at the centers of most large galaxies; they probably formed when the Milky Way and other galaxies first formed more than 13 billion years ago.

Practicing SCIENCE

What is the evidence that the center of our Galaxy contains a large, compact mass? Recall that the only direct way scientists have to measure the mass of an astronomical object is to watch something orbit around it. Then, you can use Kepler's third law to find the mass inside the orbit.

That procedure sounds simple, but, because of interstellar dust, astronomers can't see to the center of our Galaxy at visual wavelengths. However, infrared observations can detect individual stars orbiting Sgr A*. The star SO-2 has been particularly well observed, but several other stars can be followed as they orbit the center. The sizes and periods of these orbits, interpreted using Kepler's laws, reveal that Sgr A* contains at least 4 million $M_\odot$.

Now analyze a different observation to give further insight about conditions in the Milky Way Galaxy's nucleus. ***Why does strong radiation at near-infrared wavelengths imply that vast numbers of stars are crowded into the nucleus?***

15-5 Origin and Evolution of the Galaxy

Dinosaurs left behind fossilized footprints, and our Galaxy left behind clues about its youth. Astronomers have compelling evidence that nearly all of the stars in the spherical component are old and must have formed long ago when the Galaxy was very young. Much of that evidence comes from the H–R diagrams of star clusters in the halo, but some comes from comparing abundances of chemical elements in stars located in different parts of the Galaxy.

15-5a The Element-Building Process

You can recall what you have learned about stellar structure and stellar evolution in previous chapters to understand the chemical evolution of the Galaxy. In astronomy jargon, the term "**metals**" refers to all of the chemical elements heavier than helium. (Of course, that this is not what the word *metals* means to nonastronomers.) In a later chapter, you will learn about evidence that when the Universe began it contained about 90 percent hydrogen atoms, 10 percent helium atoms, and little or no metals. The first stars that formed early in the Universe's history therefore had to be nearly pure hydrogen and helium. All of the other chemical elements have been produced by nucleosynthesis, which is the process in stars that fuses hydrogen and helium to make the heavier atoms (look back to Section 12-2).

As you already know, medium-mass stars like the Sun cannot ignite carbon fusion, but during helium fusion the heat and density in their cores can trigger some nuclear reactions that cook the gas to produce small amounts of elements heavier than helium. When the aging star pushes away its surface layers to produce a planetary nebula, some of those elements are spread back into the interstellar medium where they may become part of newly forming stars and planetary systems.

The most massive stars fuse elements up to iron, and on the way cook the gas in their cores to produce small amounts of many different atoms, including sulfur and calcium. When those stars die in supernova explosions, some of those atoms are spread back into the interstellar medium, along with rarer atoms, such as gold, platinum, and uranium, produced in the supernova explosion itself. These atoms can also be incorporated eventually into new stars and planets.

Nevertheless, metals (in the astronomer's sense, including carbon, nitrogen, and oxygen) are relatively rare in the Universe. Conventionally, astronomers graph the abundance of the elements using an exponential scale, but if you replot the data using a linear scale, you can get a clearer picture of how rare these atoms are (**Figure 15-18**).

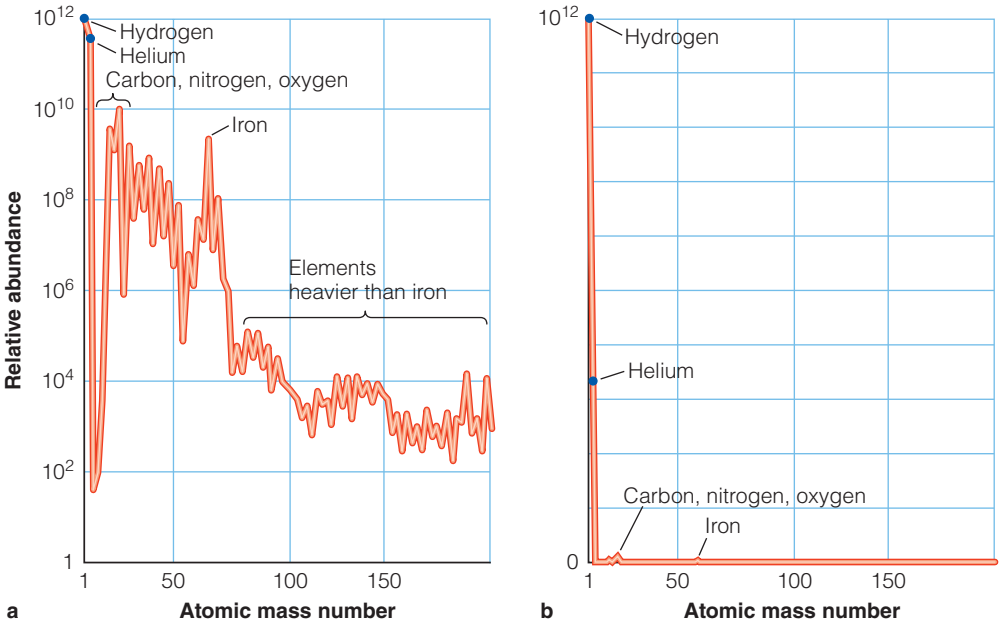
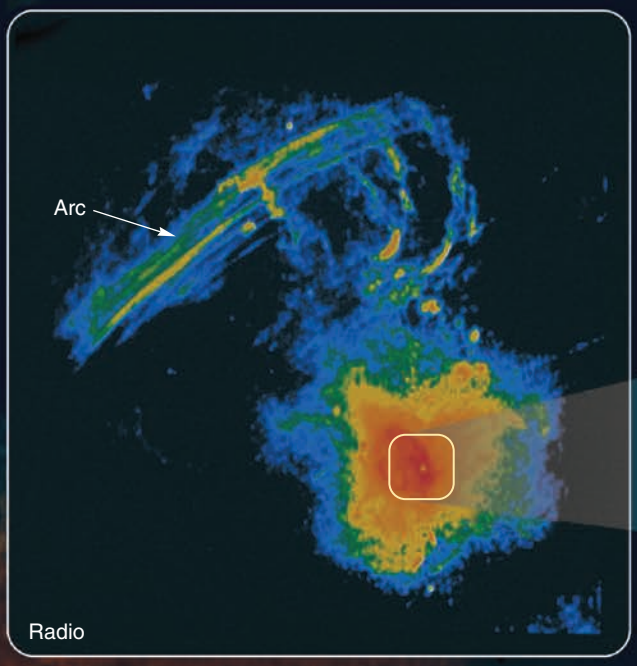


Figure 15-18 The abundances of chemical elements in the Universe. (a) When the elemental abundances are plotted on an exponential scale, you see that elements heavier than iron are about a million times less common than iron and that all elements heavier than helium (the "metals") are quite rare. (b) The same data plotted on a linear scale provide a more realistic impression of how rare the metals are. Carbon, nitrogen, and oxygen make small peaks near atomic mass 15, and iron is just visible in the graph.

Sagittarius A*

1 There is so much interstellar dust in the plane of the Milky Way that you cannot observe the nucleus of our Galaxy at visual wavelengths. The image below is a radio image of the innermost 300 pc (1000 ly). Many of the features are supernova remnants (labeled SNR), and a few are star formation clouds. Peculiar features such as threads, the Arc, and the Snake may be gas trapped in magnetic fields. At the center of the image lies the strong radio source Sagittarius A (Sgr A), the location of the nucleus of our Galaxy.



The radio map above shows Sgr A and the Arc filaments, 50 pc long. The image was made with the VLA radio telescope. The contents of the white box are shown on the opposite page.

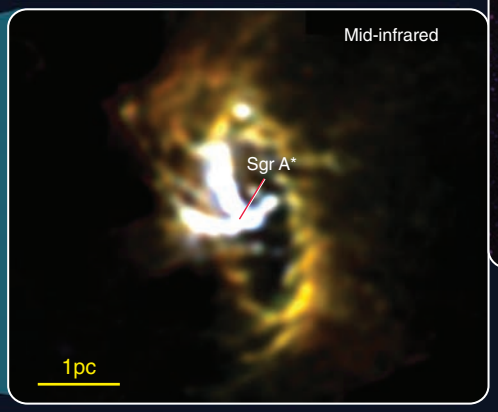
Sgr D HII
Sgr D SNR
SNR 0.9 + 0.1
Sgr B2
Sgr B1
New SNR 0.3 + 0.0
Threads
The Cane
Arc
Background galaxy
Sgr A
Threads
Radio image
NRL

Apparent angular size of the Moon for comparison

2 Infrared photons with wavelengths longer than 4 microns (4000 nm) are emitted almost entirely by warm interstellar dust. The radiation at those wavelengths coming from this part of Sagittarius is intense, indicating that this region contains lots of dust and is crowded with stars that warm the dust.

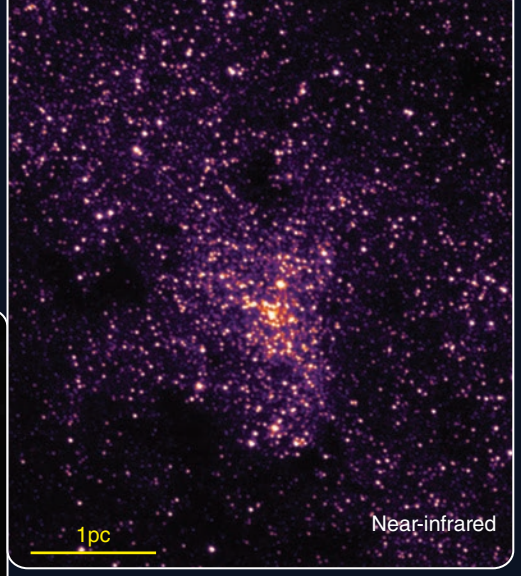
Sgr C
The Pelican
Coherent structure?
Snake
Sgr E
SNR 359.1 – 00.5

1a This high-resolution radio image of Sgr A (within the white box in the small-scale map on the opposite page) reveals a spiral swirl of gas around an intense point-like radio source known as Sgr A*. About 3 pc (10 ly) across, this spiral lies in a low-density cavity inside a larger disk of neutral gas. The arms of the spiral are thought to be streams of matter flowing into Sgr A* from the inner edge of the larger disk.



NASA/DLR/USRA/DSI/FORCAST Team/Lau et al.

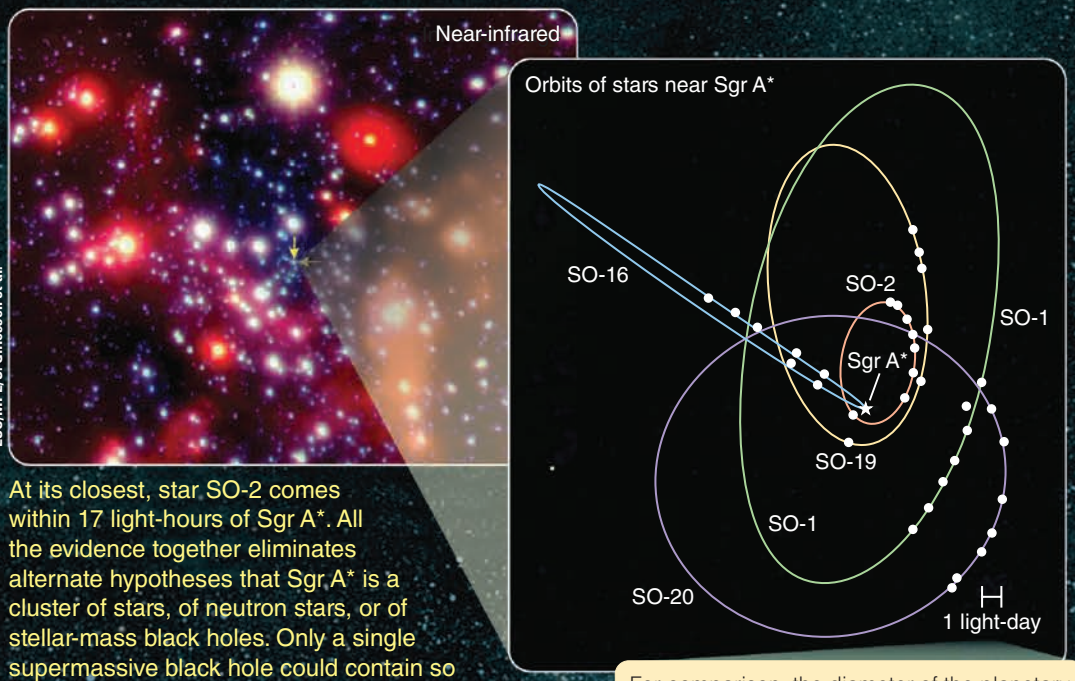
An image of the central few parsecs of our Galaxy obtained by astronomers onboard the *Stratospheric Observatory for Infrared Astronomy* (*SOFIA*), combining data at mid-infrared wavelengths of 20, 32, and 37 microns. A ring of gas and dust clouds is observed that surrounds a relatively empty central zone. The bright *y* shape represents heated material apparently falling from the cloud ring into the center. Sgr A* is located at the intersection of the arms of the *y*.



NASA/ESA/STScI/AURA/NSF/NICMOS Team

A near-infrared image from the *Hubble Space Telescope*, showing exactly the same region as the *SOFIA* mid-infrared image on the left. The dense cluster of stars located inside the circumnuclear cloud ring can be seen at this wavelength (1.9 microns), but the ring itself is too cool to emit significant amounts of near-infrared radiation. Nevertheless, especially dense parts of the ring can be discerned in this image as shadowy patches of obscuration. Sgr A* is located in the middle of the central star cluster.

3 Astronomers have been able to use large infrared telescopes and adaptive optics to follow the motions of stars orbiting around Sgr A*. A few of those orbits are shown here. The size and period of the orbit allows astronomers to calculate the mass of Sgr A* using Kepler's third law. The orbital period of the star SO-2, for example, is 15.2 years, and the semimajor axis of its orbit is 950 AU. The motions of the observed stars indicate that Sgr A* has a mass of 4 million $M_\odot$.



ESO/MPE/S. Gillessen et al.

At its closest, star SO-2 comes within 17 light-hours of Sgr A*. All the evidence together eliminates alternate hypotheses that Sgr A* is a cluster of stars, of neutron stars, or of stellar-mass black holes. Only a single supermassive black hole could contain so much mass in such a small region.

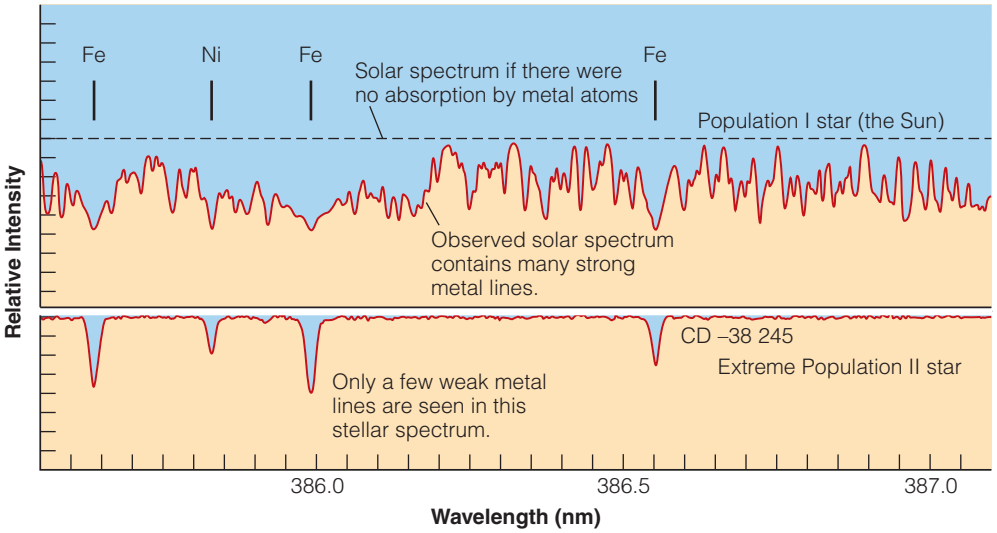
For comparison, the diameter of the planetary region of our Solar System, defined by Neptune's orbit, is half a light-day.

The *Chandra X-ray Observatory* has imaged Sgr A* and detected more than 2000 other X-ray sources in the area.

NASA/CXC/SAO/MIT/F. K. Baganoff et al.

3a A black hole with a mass of 4 million $M_\odot$ would have an event horizon with a size on the scale of this diagram smaller than the period at the end of this sentence. A slow dribble of only 0.0002 $M_\odot$ of gas per year flowing toward the black hole could produce the observed energy. A sudden increase, such as when a star falls in, could produce a violent eruption.

The evidence of a massive black hole at the center of our Galaxy seems conclusive. It is much too massive to be the remains of a dead star, however, and astronomers conclude that it probably formed as the Galaxy first took shape.



◀ **Figure 15-19** Comparison of the spectrum of the Sun, an intermediate Population I star, with that of a Population II star. (Top) Overlapping absorption lines of metals completely blanket the Population I spectrum. (Bottom) Spectrum of an extremely metal-poor star with only a few weak metal lines of iron (Fe) and nickel (Ni). This Population II star contains about 10,000 times less metal than the Sun.

15-5b Stellar Populations

By the 1940s, astronomers had accumulated evidence that there are two families of stars in the Galaxy. They form and evolve in similar ways, but they are different, especially in their abundances of metals and in their motions. **Population I stars** are metal rich, containing 2 to 3 percent metals, whereas **Population II stars** are metal poor, typically containing less than 0.1 percent metals. The difference can be seen in spectra (Figure 15-19). The abundances plotted in Figure 15-19 show metal-rich Population I composition. H–R diagrams of star clusters (Chapter 12 Concept Art: Star Clusters and Stellar Evolution) reveal that Population I stars are relatively young, and Population II stars are old.

Population I stars belong to the disk component of the Galaxy and are sometimes called disk population stars. They have nearly circular orbits in the plane of the Galaxy and formed within the past few billion years. The Sun is a Population I star, as are the type I Cepheid variables mentioned earlier in this chapter.

Population II stars belong to the spherical component of the Galaxy and are sometimes called *halo population stars*. These stars have orbits randomly inclined to the galactic plane, with shapes ranging from nearly circular to highly eccentric. They are old stars that formed when the Galaxy was young. The metal-poor globular clusters are part of the halo population, as are the type II Cepheid and RR Lyrae variable stars.

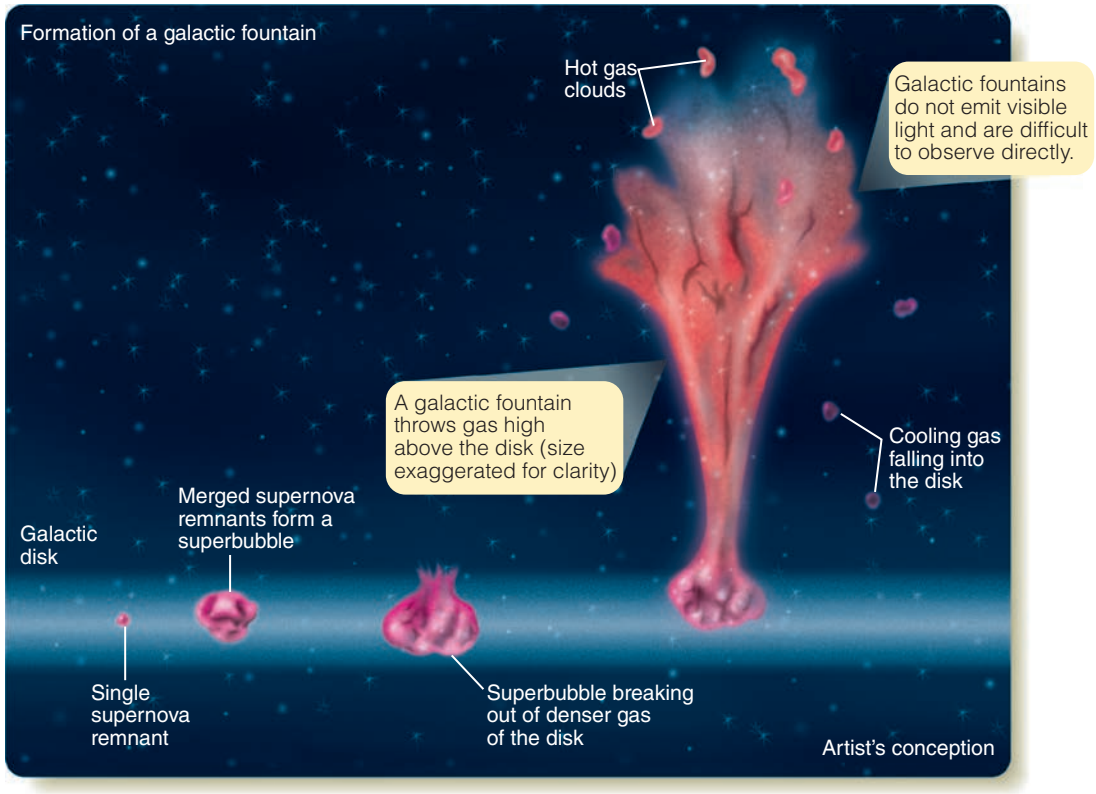
Further observations show that there is a gradation between populations. Extreme Population I stars are found only in the spiral arms. Slightly less metal-rich Population I stars, called *intermediate Population I stars*, are located throughout the disk. The Sun is an intermediate Population I star. Stars that are even less metal rich, such as stars in the central bulge, belong to the intermediate Population II. The most metal-poor stars are those in the halo, including those in globular clusters. Those are referred to as *extreme Population II stars*.

The obvious differences between the two populations of stars (Table 15-1) are clues that can help you puzzle out the history of the Milky Way Galaxy. When you observe Population II stars, such as those in the halo, you are looking at the survivors of early generations of stars in our Galaxy. The first stars evidently formed from gas that was metal poor, and the only survivors of these early generations are low-mass, long-lived stars. Their spectra still show the composition of the metal-poor gas from which they formed. (Recall that a star's spectrum shows the composition of its atmosphere, not the composition of its core that changes because of nucleosynthesis.)

Population I stars such as the Sun formed relatively recently, after the interstellar medium had been enriched in metals. Their spectra show stronger metal lines. Stars forming now have even higher metal abundances.

Table 15-1 Stellar Populations

	Population I		Population II	
	Extreme	**Intermediate**	**Intermediate**	**Extreme**
Typical location	Spiral arms; Thin disk	Thick disk	Central bulge	Halo
Metal content (%)	3	1–2	0.1–1	Less than 0.1
Orbit shape	Circular	Slightly eccentric	Moderately eccentric	Highly eccentric
Range of ages (yr)	0–0.2 billion	0.2–8 billion	8–12 billion	12–13 billion


Formation of a galactic fountain
Hot gas clouds
Galactic fountains do not emit visible light and are difficult to observe directly.
A galactic fountain throws gas high above the disk (size exaggerated for clarity)
Cooling gas falling into the disk
Merged supernova remnants form a superbubble
Galactic disk
Single supernova remnant
Superbubble breaking out of denser gas of the disk
Artist's conception

▲**Figure 15-20** In this edge-on view of the galactic disk, you can see that superbubbles, formed of multiple supernova remnants, can be roughly as large as the thickness of the disk. Once a superbubble breaks out of the denser gas of the disk, it is thought to produce a galactic fountain. The metal-enriched gas could flow out into the halo where it would cool and fall back to enrich the interstellar medium in metals.

15-5c Galactic Fountains

Supernova remnants are rich in metals and eventually mix back into the interstellar medium. In fact, supernovae continuously stir and enrich the interstellar medium. But a larger-scale process may be even more efficient at spreading newly created metals throughout the disk of the Galaxy, where they can be incorporated into newly forming stars.

In Section 10-2d, you learned that neighboring supernova remnants can merge to form a superbubble of hot gas. A supernova remnant may be only a few tens of parsecs in diameter, but a superbubble can be more than ten times bigger. The denser gas of the galactic disk tends to confine an expanding superbubble, but the disk is only a few hundred parsecs thick. If a superbubble breaks out of the denser gas of the galactic disk, it can spew hot, metal-enriched gas high above the plane of the Galaxy in a **galactic fountain** (Figure 15-20). As this gas cools and falls back into the disk, it can spread metals through the Galaxy.

Galactic fountains are difficult to observe directly, but isolated clouds of gas have been found high in the halo, and cool clouds have been found falling toward the disk. Some of these may be material from outside the Galaxy that is falling inward for the first time, but the lower-velocity clouds only a few thousand parsecs above the disk may be cooling gas that was spewed out of the disk by fountains. It seems likely that this process contributes to spreading metals through the disk. Anyway, it is evident that supernovae create metals and mix them back into the interstellar medium where they can be incorporated into newly forming stars.

Because the recycling of stellar nucleosynthesis products means that the metal abundance of newborn stars increases as the Galaxy ages, astronomers can use metal abundance as an index to the ages of the components of our Galaxy. The halo is old, and the disk is clearly younger.

15-5d Age of the Galaxy

Because astronomers know how to find the age of star clusters, they can estimate the age of the oldest stars in the Galaxy, giving a lower limit to the age of the entire Galaxy. The process sounds straightforward, but uncertainties make the answer hard to interpret.

The oldest open clusters are 9 billion to 10 billion years old. These ages come from the turnoff points in their H–R diagrams, but finding the age of an old cluster is difficult because old clusters change so slowly. Also, the exact location of the turnoff point depends on chemical composition, which differs slightly among clusters. Finally, open clusters are not strongly bound by gravity, so older open clusters may have dissipated as their stars wandered

away. It is reasonable to suppose that the galactic disk is somewhat older than the oldest remaining open clusters, which suggests that the disk is at least 10 billion years old.

Globular clusters have low-luminosity turnoff points in their H–R diagrams and are clearly old, but finding these ages is difficult. For globular clusters as for open clusters, slight differences in chemical composition make noticeable differences when calculating the stellar models from which ages are determined. Also, to find the age of a cluster, astronomers must know its distance. Parallaxes from the *Hipparcos* satellite have allowed astronomers to improve the calibration of the Cepheid and RR Lyrae variable stars, and careful studies with the newest large telescopes have refined the chemical compositions and better defined the H–R diagrams of globular clusters. An analysis of all the data suggests that the average globular cluster's age is about 11 billion years, although some globular clusters are younger than that, and some are older. Studies of the oldest globular clusters suggest that the halo of our Galaxy is at least 13 billion years old.

Observations of stellar populations and clusters show that the disk is younger than the halo. You can combine these ages with the process of nucleosynthesis to tell the story of our Galaxy.

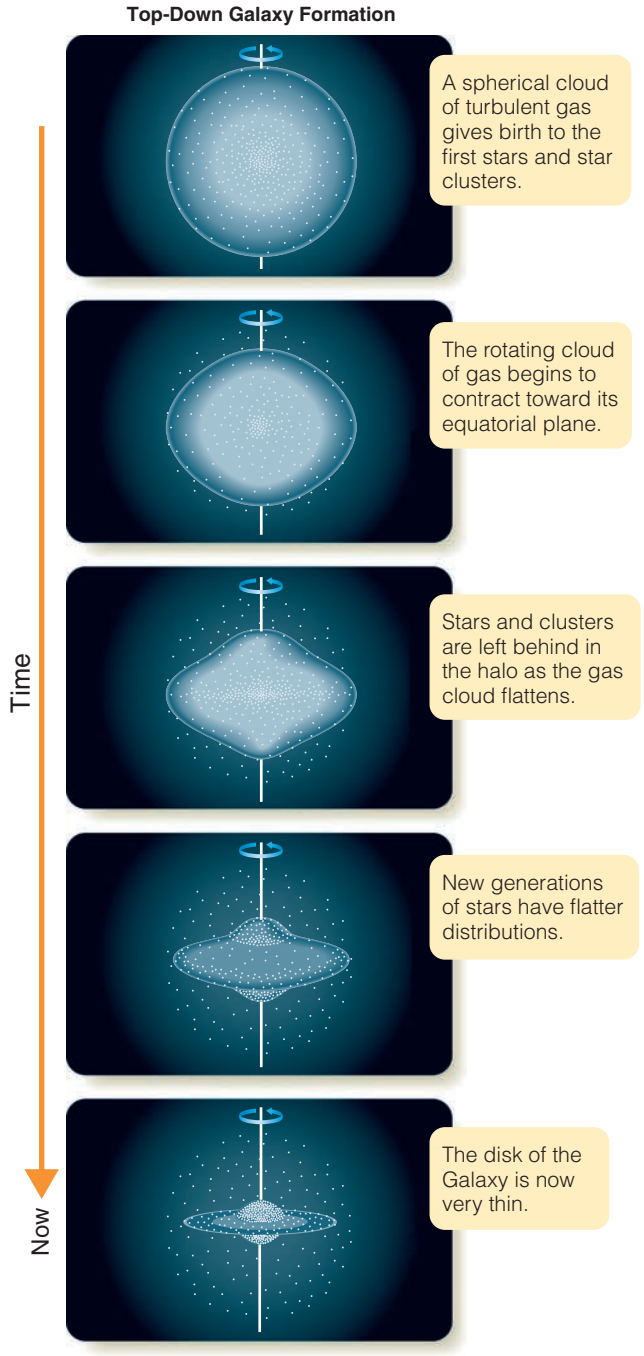
15-5e Formation of the Galaxy

Beginning in the 1950s, astronomers developed the **top-down hypothesis**, sometimes called the *monolithic collapse hypothesis* to explain the formation of our Galaxy that was based on the data available at the time (Figure 15-21). Later observations forced a reevaluation and modification of that hypothesis.

The lack of metals in the spherical component of the Galaxy tells you it is very old, a fossil left behind by the Galaxy when it was young and dramatically different from its present disk shape. The top-down hypothesis proposed that the Galaxy formed from a single large **protogalaxy** cloud of turbulent gas more than 13 billion years ago. That cloud contracted to form our Galaxy.

As gravity pulled the gas inward, turbulence in the parent cloud caused it to fragment into smaller clouds with random relative velocities. As a result, the stars and star clusters that formed from the cloud fragments had orbits with a wide range of shapes; a few were circular, but most were eccentric. The orbits were also inclined to the galactic plane at different angles, resulting in a spherical cloud of stars—the spherical component of the Galaxy. Those first stars would have been metal-poor because no stars had existed earlier to enrich the gas with metals.

The second stage in this hypothetical history accounts for the formation of the disk component. As gas clouds in the original spherical cloud collided, turbulent motions in the gas canceled out, like swirls in recently stirred coffee, and the cloud was left with an average, uniform rotation. But a rotating, low-density cloud of gas cannot remain spherical because the internal pressure is too low to support the weight. Such a cloud will collapse and spin faster because of conservation of angular momentum. The combination of collapse plus increasing spin rate would cause the cloud to flatten (Figure 15-21). Eventually, according to the top-down hypothesis, this process would have produced the Galaxy's disk.


Top-Down Galaxy Formation
A spherical cloud of turbulent gas gives birth to the first stars and star clusters.
The rotating cloud of gas begins to contract toward its equatorial plane.
Stars and clusters are left behind in the halo as the gas cloud flattens.
New generations of stars have flatter distributions.
The disk of the Galaxy is now very thin.
Time
Now

▲ **Figure 15-21** The top-down model for the origin of our Galaxy begins with a spherical gas cloud that flattens into a disk.

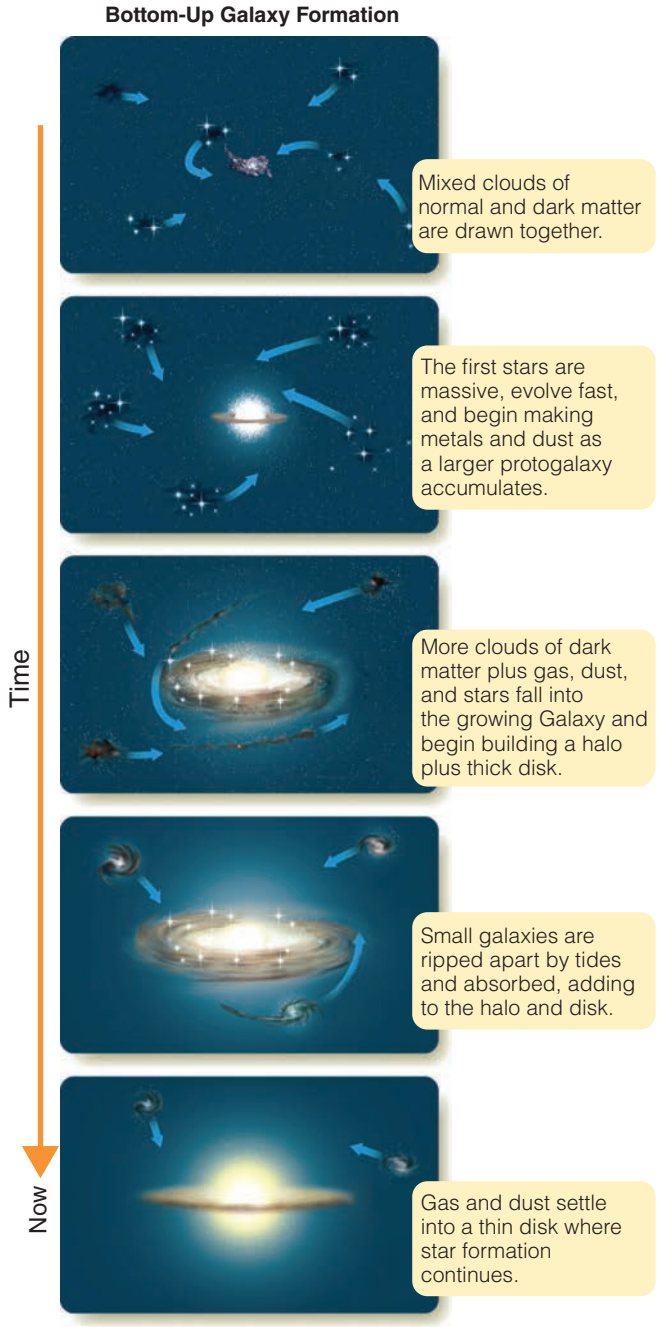
This contraction from a sphere into a disk would have taken billions of years, and while that happened the metal abundance would have gradually increased as generations of stars were born from the gradually flattening gas cloud. The stars and globular clusters that had already formed in the halo were left behind by the gas cloud as it flattened, so subsequent generations of stars formed in flatter and flatter distributions. The gas distribution in the Galaxy now is so flat that the youngest stars are confined to a disk only about 100 pc thick. These stars are metal rich and have nearly circular orbits.

The top-down hypothesis accounts for many of the Milky Way's properties. However, larger telescopes and more sensitive instruments allowed more precise measurements of motions and metallicities. Eventually, contradictions between predictions of the top-down hypothesis and newer observations were noticed. For example, the top-down hypothesis predicts that the globular star clusters in the halo should be all about the same age, with the most distant ones being slightly older than the rest. Instead, globular clusters have a surprisingly wide range of ages, and some of the youngest clusters are in the outer part of the halo.

The answer to the puzzle of location versus age of globular clusters may lie in a major modification of the top-down, monolithic-collapse hypothesis. The central bulge and some of the halo might in fact have formed by monolithic collapse of a single gas cloud. But then, substantial amounts of new material seems to have been added to the young Milky Way Galaxy by collisions with already-formed star clusters and fresh gas clouds falling in from intergalactic space. As a result, the Galaxy today would have components with a range of ages and motions representing the pieces from which it was assembled. Evidence supporting this hypothesis is the discovery of several streams and rings of stars surrounding our Galaxy, each with a unique age and metallicity. Astronomers hypothesize that these star streams were produced when several small galaxies were captured, tidally ripped apart, and absorbed by our Galaxy. You will see clear evidence in the next chapter that such galaxy mergers do occur.

Thus, the new **bottom-up hypothesis** is that the Milky Way Galaxy was partly assembled from smaller units, absorbing several smaller galaxies that had already partially evolved plus a number of infalling gas clouds (**Figure 15-22**). That hypothesis could explain the observed ranges of globular cluster ages, metallicities, and locations. In this scenario, some of the Galaxy's globular clusters are hitchhikers from other galaxies that joined to make our Galaxy.

Another problem with the top-down hypothesis is that, although the oldest stars in the Galaxy are metal poor, they still have some metals. There must have been at least a few massive stars to create those metals during a generation before the formation of the oldest stars now seen in the halo.


Bottom-Up Galaxy Formation
Time
Now
Mixed clouds of normal and dark matter are drawn together.
The first stars are massive, evolve fast, and begin making metals and dust as a larger protogalaxy accumulates.
More clouds of dark matter plus gas, dust, and stars fall into the growing Galaxy and begin building a halo plus thick disk.
Small galaxies are ripped apart by tides and absorbed, adding to the halo and disk.
Gas and dust settle into a thin disk where star formation continues.

▲**Figure 15-22** Artist's conception: The bottom-up hypothesis for the formation of the Milky Way Galaxy proposes that smaller star systems accumulated to form larger ones. To see how this could have built our Galaxy, start with the first frame, only a few hundred million years after the beginning of the Universe, as small clouds of matter begin accumulating and stars begin forming in them. In the second frame from the top the central object has grown larger, and in the third frame the galactic halo and disk are forming. By the time shown in the bottom frame, representing today, the disk of the Galaxy has become very thin.

The answer to the puzzling fact that the oldest stars have some metals may lie in a unique era of star formation early in the Galaxy's history. As you will learn in a later chapter, astronomers have evidence that the protogalaxy cloud should

have been purely hydrogen and helium with almost no metals (plus dark matter that would not have affected these processes). Models of star formation from metal-free material indicate that the stars would all have been very massive, so this first generation of stars would have evolved rapidly, exploded as supernovae, and enriched the protogalaxy with traces of metals. None of those massive zero-metal first-generation stars would survive to the present, but the metals they created are detectable in the oldest Population II stars.

The metal abundances and ages of stars and star clusters in our Galaxy are important clues about its history, but metal abundance and age probably do not tell the whole story about how the Milky Way Galaxy compares with other galaxies. Astronomer Bernard Pagel was thinking of this when he said, "Cats and dogs may have the same age and metallicity, but they are still cats and dogs."

Practicing SCIENCE

Why do metal-poor stars have the most eccentric orbits?

A scientist needs to be able to spot connections between seemingly unrelated facts to gain understanding of the "big picture."

The composition and motion of a star would seem to be unrelated. Certainly, the metal abundance of a star cannot affect its orbit, so a scientist must be careful not to confuse cause and effect with the relationship between these two facts. In other words, if A and B go together, it doesn't necessarily mean that A causes B or that B causes A. A and B might both be caused by something else, C.

Astronomers hypothesize that chemical composition and orbital shape of stars in the Milky Way depend on a third factor—age. The oldest stars are metal poor because they formed before there had been many supernova explosions to create and scatter metals into the interstellar medium. Those stars formed long ago when the Milky Way Galaxy was young and gas motions were turbulent and not yet organized into a disk. As a result the stars tended to take up randomly shaped orbits, many of which are quite elongated. Consequently, today, the most metal-poor stars tend to follow the most eccentric orbits.

Nevertheless, even the oldest stars known in our Galaxy contain some metals. They are metal poor, but not completely metal free. Ponder another important question about the Galaxy's history: ***Where did those metal-poor stars get their metals?***

WHAT ARE WE?

Children of the Milky Way

Hang on tight. The Sun, with Earth in its clutch, is ripping along at about 225 km/s (that's 500,000 mph) as it orbits the center of the Milky Way Galaxy. We live on a wildly moving ball of rock in a large galaxy that humanity calls home, but the Milky Way is more than just our home. Perhaps "parent galaxy" would be a better name.

Except for hydrogen atoms, which have survived unchanged since the Universe began, you and Earth are made of metals—atoms heavier than helium. There is no helium in your body, but there is plenty of carbon, nitrogen, and oxygen. There is calcium in your bones and iron in your blood. All of those atoms and more were cooked up inside stars or during their supernova death throes.

Stars are born when clouds of gas orbiting the center of our Galaxy collide with the gas in spiral arms and are compressed. That process has given birth to generations of stars, and each generation has produced elements heavier than helium and spread them back into the interstellar medium. The abundance of metals has grown slowly in our Galaxy. About 4.6 billion years ago, a cloud of gas enriched in those heavy atoms slammed into a spiral arm and produced the Sun, Earth, and, eventually, you. You have been created by the Milky Way—your parent galaxy.

Study and Review

SUMMARY

15-1 Discovery of the Galaxy

- William and Caroline Herschel counted stars at many locations over the sky to show that our star system seemed to be shaped like a grindstone, with the Sun near the center. Later astronomers incorrectly concluded the galaxy is only about 10 **kiloparsecs (kpc)** in diameter with the Sun at the center.
- Using the proper motions of a few Cepheids, Shapley was able to **calibrate** the period–luminosity relation discovered by Henrietta Leavitt and thus estimate the distance to globular clusters. His work demonstrated that our Galaxy is much larger than the portion that can easily be observed from Earth and that the Sun is not at the center.

15-2 Structure of the Galaxy

- Modern observations suggest our Galaxy contains a **disk component** about 80,000 ly in diameter. The Sun is located about two-thirds the distance from the center to the visible edge. The **central bulge** and an extensive **halo**, which contains old stars and little gas and dust, make up the **spherical component**.

- The mass of the Milky Way Galaxy can be found from the Galaxy's **rotation curve**. Kepler's third law reveals that the Milky Way Galaxy contains at least 100 billion $M_\odot$. If stars orbited in **Keplerian motion**, more distant stars would orbit more slowly. As observations show that they do not, the halo must contain much more mass than is visible. Because the mass in this **galactic corona** is not emitting detectable electromagnetic radiation, astronomers call this mass **dark matter**.

15-3 Spiral Arms and Star Formation

- You can trace the spiral arms within the disk of the Milky Way Galaxy through the Sun's neighborhood by using **spiral tracers** such as OB associations. O and B stars live short lives and don't have time to move far from their place of birth. Because they are found mostly scattered along the leading edges of spiral arms, astronomers conclude that spiral arms are sites of star formation.
- To extend the map over the entire Milky Way Galaxy, astronomers must use radio, infrared, and X-ray telescopes to see through the dust in the interstellar medium.
- The **spiral density wave theory** describes spiral arms as regions of compression that orbit around the center of a galaxy within the disk. When a faster orbiting gas cloud overtakes a slower orbiting compression wave, the gas cloud compresses, stimulating star formation. A **grand-design galaxy** with its two-armed spiral visibly highlights the two density waves. The Milky Way Galaxy is in between the extremes of grand design and **flocculent galaxy** spiral patterns.
- Another process, **self-sustaining star formation**, may act to modify the arms by producing branches and spurs. These offshoots may be generated by the birth of massive stars, which triggers the formation of more stars by compressing neighboring gas clouds.

15-4 The Nucleus of the Galaxy

- The nucleus of our Galaxy is invisible at visual wavelengths. However, radio, infrared, and X-ray radiation can penetrate the gas and dust. These wavelengths reveal a crowded field of stars and warm dust.
- The very center of the Milky Way Galaxy is marked by a unique radio source, Sagittarius A* (Sgr A*). This gravitational core must be less than 1 AU in diameter and the motions of stars around the center show that it must contain about 4 million $M_\odot$. A supermassive black hole is the only object that could contain so much mass in such a small space.

15-5 Origin and Evolution of the Galaxy

- Stars manufacture elements heavier than helium, called **metals** by astronomers, by nucleosynthesis processes in their cores; as stars die, they spread these metals back into the interstellar medium.
- Metals are relatively rare in the Universe. **Population I stars**, including the Sun, are richer in metals than **Population II stars**. This shows that Population I stars formed more recently, after the interstellar medium had been enriched with metals from a previous generation of stars.
- **Galactic fountains**, produced by expanding supernova remnants, may help spread metals throughout the disk.
- Using metal abundance, astronomers conclude that the halo formed first and the disk later. The oldest star clusters reveal that the disk of our Galaxy is younger than the halo. Because the oldest globular clusters appear to be about 13 billion years old, our Galaxy must have formed about 13 billion years ago.
- The **top-down hypothesis** suggests that the Milky Way Galaxy formed from a single, roughly spherical **protogalaxy** cloud of gas that gradually flattened into a disk. This hypothesis is not supported by some of the current evidence. A newer, **bottom-up hypothesis** includes mergers of the young Milky Way Galaxy with many separately formed smaller galaxies and star clusters plus clouds of infalling gas.

REVIEW QUESTIONS

1. What evidence can you give that we live in a galaxy?
2. Why is it difficult to give exact dimensions for the Galaxy's disk and visible halo?
3. Why didn't astronomers before Shapley realize how large our Galaxy is?
4. What evidence can you give that our Galaxy has a galactic corona?
5. What color does the galactic corona have, if any? Why?
6. Contrast the motion of the disk stars to that of the halo stars. How do their orbits differ and why?
7. Which parts of a spiral galaxy comprise the spherical component?
8. What evidence can you give that we live in a spiral galaxy?
9. If you could look at the Milky Way Galaxy edge-on in visual wavelengths, would you see the spiral nature of the arms?
10. Why can't spiral arms be physically connected structures? What would happen to them if they were physically connected structures?
11. List three different spiral tracers. Why is each one a tracer of spiral structure?
12. Why do all spiral tracers seem to be young?
13. Do the stars within the spiral arms, and the spiral arms, orbit the center of the Galaxy at the same velocity? Why or why not?
14. Are spiral arms the same as density waves? Why or why not?
15. What color is the central bulge? The spiral arms? Why?
16. If a galaxy's only process producing spiral patterns were the spiral density wave, which type of spiral galaxy would that be? Why?
17. Is the Milky Way Galaxy a grand-design spiral, a flocculent spiral, or in between? How do you know?
18. How does self-sustaining star formation produce clouds of stars that look like segments of spiral arms?
19. Why must astronomers use infrared and radio telescopes to observe the motions of stars around Sgr A*?
20. What evidence can you give that the nucleus of our Galaxy contains a supermassive black hole?
21. Could the black hole in the nucleus of the Milky Way Galaxy be the remnant of a single dead star? Why or why not?
22. Which formed last: the visible halo, the galactic nucleus, or the disk? How do you know?
23. Why are metals less abundant in older stars than in younger stars?
24. Where are the stars with the most abundant metals in the Milky Way Galaxy: the visible halo, the disk, the spiral arms, or in the bulge? Why?
25. Why do metal-poor stars have a wider range of orbital shapes than metal-rich stars like the Sun?
26. How does a galactic fountain spread metals through the Galaxy's disk?
27. Rank these objects from oldest to youngest: the Solar System, the Universe, the Milky Way Galaxy.
28. What evidence contradicts the top-down hypothesis for the origin of our Galaxy?
29. **How Do We Know?** Calibration simplifies complex measurements. How is the work of later astronomers dependent on the expertise of the astronomer who did the initial calibration?
30. **How Do We Know?** The story of a process makes the facts easier to remember, but easy recall is not the true goal of the scientist. What is the real value of understanding a scientific process?

ACTIVE INQUIRY QUESTIONS

1. How would the information in this chapter differ if interstellar dust did not block starlight?
2. The three most important attributes in real estate are said to be location, location, and location. If you could relocate our Solar System to another region in our Galaxy, where would that be? Why?

PROBLEMS

1. Make a scale sketch of the Milky Way Galaxy in cross section (that is, edge-on). Include the disk, Sun, nucleus, visible halo, and some globular clusters. Try to draw the globular clusters to scale.
2. Because of interstellar dust, astronomers can see at most about 5 kpc into the disk of the Milky Way Galaxy at visual wavelengths. What percentage of the galactic disk's area does that include? (*Hint:* Consider the area of the entire disk versus the area visible from Earth.)
3. If a standard passenger aircraft can fly at 0.25 km/s (560 mph), how long would it take to reach the Sun? The galactic center? (*Notes:* The distance to the Sun can be found in Chapter 8 Celestial Profile: The Sun, 1 pc $= 3.1 \times 10^{13}$ km, and 1 yr $= 3.2 \times 10^{7}$ s.)
4. If a typical halo star has an orbital velocity of 250 km/s, how many years does it take to pass through the disk of the Milky Way Galaxy? Assume the disk is 1000 pc thick and the star passes perpendicularly through it. (*Notes:* 1 pc $= 3.1 \times 10^{13}$ km and 1 yr $= 3.2 \times 10^{7}$ s.)
5. Find the distance to Polaris, a type I (classical) Cepheid, using its 3.97-day period. (*Note:* Polaris's apparent visual magnitude is +2.0.) (*Hints:* See Figure 12-14, and use the magnitude–distance formula, Eq. 9-2b.)
6. If the RR Lyrae stars in a globular cluster have average apparent visual magnitudes of +14.0, how far away is the cluster? (*Hints:* See Figure 12-14, and use the magnitude–distance formula, Eq. 9-2b.)
7. If interstellar dust makes an RR Lyrae variable star look 1 magnitude fainter than the star should, by how much will you over- or underestimate its distance? Is your calculated value an over- or underestimate? (*Hint:* Use the magnitude–distance formula, Eq. 9-2b.)
8. If you assume that a globular cluster 4.0 arc minutes in diameter is actually 25 pc in diameter, how far away is it? (*Hint:* Use the small-angle formula, Eq. 3-1.)
9. If the Sun is 4.6 billion years old, how many times has it orbited the Milky Way Galaxy?
10. If astronomers find they have made a mistake and our Solar System is actually 7.3 rather than 8.3 kpc from the center of the Milky Way Galaxy but the orbital velocity of the Sun is still 225 km/s, what is the minimum mass of the Galaxy within the orbit of the Sun? (*Hints:* Review the calculation in Section 15-2b, and use Kepler's third law, Eq. 4-1.)
11. What temperature would a cloud of interstellar dust have to be to radiate most strongly at a wavelength of 100 μm? (*Hint:* Use Wien's law, Eq. 7-2.) (*Note:* 1 μm = 1000 nm.)
12. Infrared radiation from the center of our Galaxy with a wavelength of about 2 μm (2×10^{-6} m) comes mainly from cool stars. What is the surface temperature of those stars? (*Hint:* Use Wien's law, Eq. 7-2.) (*Note:* 1 μm = 1000 nm.)
13. If an object at the center of the Milky Way Galaxy has a linear diameter of 1.0 AU, what is its angular diameter as seen from Earth? Assume the distance to the center of the Galaxy is 8.3 kpc. (*Hint:* Use the small-angle formula, Eq. 3-1.)
14. The Sun's absolute magnitude is +4.8. Using the magnitude–distance formula (Eq. 9-2), what would be the apparent magnitude of the Sun viewed from the distance of the galactic center, 8.3 kpc, assuming no extinction? Could the *Hubble Space Telescope* see something that faint? (*Hint:* Refer to Figure 2-6.) Now include 30 magnitudes of extinction from interstellar dust. What would the Sun's apparent magnitude be? Could the *Hubble Space Telescope* see it in this case?
15. What is the ratio of metals in Population I stars to that in Population II stars? Is this difference significant or insignificant?

SENSE OF PROPORTION

1. The Milky Way Galaxy is about 100 million times larger in diameter than the planetary region of our Solar System (i.e., the diameter of Neptune's orbit). If you represent our Solar System with a poker chip having a diameter of about 1.5 inches, how big a disk would you need to represent our Galaxy? Express your answer in miles.
2. If you represent the age of the Universe, 13.8 billion years, by the length of an American football field, 100 yards = 300 feet, what length would represent the time for the Sun to orbit the Milky Way Galaxy once, 225 million years?

LEARNING TO LOOK

1. Look at the star charts in Appendix B. If you were located in M31, the Andromeda Galaxy, and could see the Milky Way Galaxy, what would we look like? Would you see the Milky Way Galaxy edge-on, face-on, or some angle in between? How do you know?
2. Look at Figure 15-3. Why is the center of the Milky Way Galaxy in a dark portion of the sky map?
3. Look at Figure 15-6. Compare and contrast the two false-color images of the Milky Way Galaxy. Where is the cold dust located? The warm dust? Where can we say there is probably no dust?
4. Why does the galaxy shown here have so much dust in its disk? How big do you suppose the halo of that galaxy really is, relative to what you can see in this picture?

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

5. Why are the spiral arms in the galaxy in this image blue? What color would the halo be if the halo were bright enough to be seen in this photo?

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/A. Reiss

16 Galaxies

NASA/ESA/STScI/AURA/NSF/
Hubble Heritage Team

▲ Galaxy M104, nicknamed The Sombrero, is relatively nearby, at a distance of only 31 million light-years. Like our Milky Way Galaxy, it is a spiral galaxy with abundant clouds of gas and dust in its disk plane where stars are forming. Also like the Milky Way, M104 has a supermassive black hole in its nucleus.

Our Milky Way Galaxy is only one of the many billions of galaxies visible in the sky. This chapter will expand your horizon to view the different kinds of galaxies and their complex histories. Here you can expect answers to five important questions:

- What are the different types of galaxies?
- How do astronomers measure the distances to galaxies?
- How do galaxies differ in size, luminosity, and mass?
- Do other galaxies contain supermassive black holes and dark matter, as does our Galaxy?
- Why are there different kinds of galaxies?

As you begin studying galaxies, you will discover they are classified into different types, and that will lead you to insights into how galaxies form and evolve. In the next chapter, you will discover that some galaxies are violently active, and that will give you more clues to the history of galaxies and of the Universe.

> **A hypothesis or theory is clear, decisive, and positive, but it is believed by no one but the man who created it. Experimental findings, on the other hand, are messy, inexact things which are believed by everyone except the man who did that work.**
>
> HARLOW SHAPLEY, *THROUGH RUGGED WAYS TO THE STARS*

SCIENCE FICTION HEROES FLIT EFFORTLESSLY among the stars, but almost none travels between the galaxies. As you leave your home galaxy—the Milky Way—behind, you will voyage out into the depths of the Universe, out among the galaxies, into space so deep it is unexplored even in fiction.

Before you can begin to understand the life stories of the galaxies, you need to gather some basic data. How many kinds of galaxies are there? How big are they? How massive are they? You can characterize the family of galaxies just as you characterized the family of stars in Chapter 9.

16-1 The Family of Galaxies

Less than a century ago, astronomers did not know for certain that there are such things as galaxies. Nineteenth-century telescopes revealed faint nebulae scattered among the stars, and some were spiral shaped. Astronomers argued about the nature of these faint nebulae, but it was not until the 1920s that the understanding was reached that some of those nebulae were entire other galaxies outside our own. Only in recent decades have astronomical telescopes fully revealed their tremendous beauty and intricacy (**Figure 16-1**).

NGC 3341a 117 million ly
NGC 3341b 140 million ly
Clouds of gas and dust in a small foreground galaxy visible in silhouette against a background galaxy.
Visual
NGC 4414 60 million ly
Young blue stars illuminate spiral arms.
Visual
ESO 510-G13
150 million ly
Dusty disk of galaxy warped by interaction with another galaxy
New stars forming in dust clouds
Visual
M83 12 million ly
Dust clouds glow red in this infrared image.
Infrared
NGC 4414, ESO 510-G13, NGC 3341a and 3341b: NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team M83: ESO

▲ **Figure 16-1** A century ago, photos of galaxies looked like clouds of haze. Modern images of these relatively nearby spiral galaxies reveal impressively beautiful objects filled with stars plus clouds of gas and dust.

16-1a The Discovery of Galaxies

Galaxies are faint objects; only one, Messier 31, is visible to the unaided eye (look back to Section 15-1a). So galaxies were not noticed until telescopes had grown large enough to gather significant amounts of light. In 1845, William Parsons, third Earl of Rosse in Ireland, built a telescope 72 inches (1.8 meters) in diameter. It was, for a span of more than 70 years, the largest telescope in the world. Parsons lived before the invention of astronomical photography, so he had to view the faint nebulae directly at the eyepiece and sketch their shapes. He noticed that some have a spiral shape, and they became known as **spiral nebulae**. Parsons concluded that the spiral nebulae were great spiral clouds of stars. In the 18th century the German philosopher Immanuel Kant had proposed that the Universe was filled with great wheels of stars, which he called *island universes*. Parsons adopted that term for the spiral nebulae.

Not everyone agreed that the spiral nebulae were clouds of stars lying outside our own star system. An alternate point of view was that the Milky Way star system was alone in an otherwise empty Universe. Sir William Herschel had counted stars in different directions and concluded that the star system was shaped approximately like a grindstone (look back to Figure 15-2). Space was imagined to be a limitless void beyond the edge of this grindstone star system. In this view, the spiral nebulae were nothing more than whirls of gas and faint stars within our star system.

The debate over the spiral nebulae could not be resolved in the 19th century because the telescopes were not large enough and photographic plates, when they became available in the late 1800s, were not sensitive enough. The spiral nebulae remained foggy swirls even in the best photographs, and astronomers continued to disagree over their true nature. In April 1920, two astronomers debated the issue at the National Academy of Sciences in Washington, D.C. Harlow Shapley of Mount Wilson Observatory, whom you met in the previous chapter, had recently shown that the Milky Way was a much larger star system than had been thought. He argued that the spiral nebulae were just objects within the Milky Way star system. Heber Curtis of Lick Observatory argued that the spiral nebulae were island universes far outside the Milky Way star system.

Historians of science mark this **Shapley–Curtis debate** as a turning point in modern astronomy, but at the time it was inconclusive. To make their arguments, Shapley and Curtis each cited items of evidence that astronomers now know were incorrect. The disagreement was finally resolved, as is often the case in astronomy, by observations with a bigger telescope.

At the end of 1924, Mount Wilson astronomer Edwin Hubble (for whom the *Hubble Space Telescope* is named) announced that he had taken photographic plates of a few bright spiral nebulae using the new 100-inch Hale telescope. He had detected individual stars in the spiral nebulae and identified some of those stars as Cepheid variables with apparent magnitudes of about +18 but month-long pulsation periods, indicating they are supergiants (look back to Figure 12-14). To appear that faint, those luminous Cepheids had to be very distant; therefore the spiral nebulae must be outside the Milky Way star system. They are galaxies, separate from ours, and rivaling ours in size.

16-1b How Many Galaxies Are There?

Like leaves on the forest floor, galaxies carpet the sky. Pick any direction in the sky away from the obscuring dust and gas of the Milky Way, and you are looking deep into space. Photons that have traveled for billions of years enter your eye, but they are too few to register on your retina. Only the largest telescopes can gather enough light to detect distant galaxies.

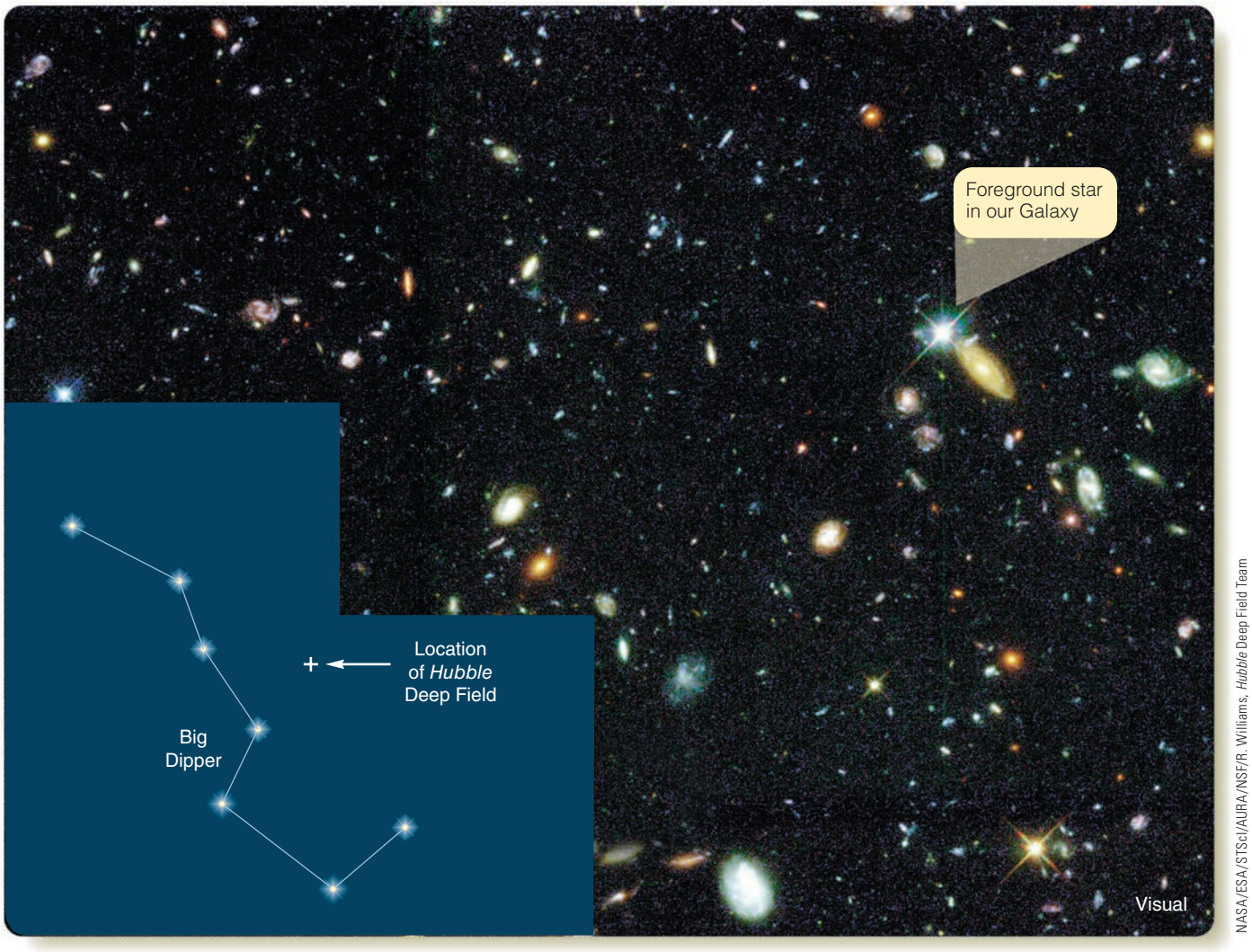
When astronomers chose a relatively empty spot on the sky near the Big Dipper and used the *Hubble Space Telescope* to record a 10-day exposure, they found thousands of galaxies crowded into the image. That image, now known as the Hubble Deep Field, contains a few relatively nearby galaxies along with many others that are as much as 10 billion light-years away (**Figure 16-2**).

Since the first Hubble Deep Field was recorded, other deep fields have been imaged in other parts of the sky and at other wavelengths. The Great Observatories Origins Deep Survey (GOODS) program used the *Hubble Space Telescope* with other space observatories including the infrared *Spitzer Space Telescope* and the *Chandra X-ray Observatory* plus some of the largest ground-based telescopes to image more deep fields. All of these deep-field images show the entire sky to be thickly covered with galaxies.

These surveys of distant galaxies are especially amazing when you think that, less than a century ago, humanity did not know that we live in a galaxy or that the Universe contained other galaxies. Today telescopes are capable of detecting a few hundred billion galaxies at a wide range of wavelengths, and new, larger telescopes will reveal even more. The discovery of galaxies was one of the most important turning points in the history of astronomy and in the history of the quest to understand humanity's true location in the Universe.

16-1c The Shapes of Galaxies

Look at the galaxies in the Hubble Deep Field (Figure 16-2), and you will see a variety of shapes. Although many of the most prominent ones are spiral, others are elliptical, and some are irregular. Astronomers classify galaxies according to their shapes in photographs made at visual wavelengths using a system developed by Edwin Hubble in the 1920s. Such systems of classification are a fundamental technique in science (**How Do We Know? 16-1**).



▲ **Figure 16-2** Seen with the naked eye, the area shown in this image would be an apparently empty spot on the sky only one-thirtieth the diameter of the full moon. In fact, this extremely long time exposure, known as the *Hubble* Deep Field North, contains more than 1500 galaxies. Only four stars are visible in this image; they are sharp points of light with diffraction spikes produced by the telescope optics. Evidently the entire sky is similarly filled with galaxies.

Study **Galaxy Classification** and notice three important points and five new terms that describe the main types of galaxies:

1. Many galaxies have no disk, no spiral arms, and almost no gas and dust. These *elliptical galaxies* (class E) range from huge giants to small dwarfs. They generally have a red tint, signifying a lack of massive young stars.
2. Disk-shaped galaxies usually have spiral arms, contain gas and dust, and have a blue tint. Some of these *spiral galaxies* (class S) have a central region shaped like an elongated bar and are called *barred spiral galaxies* (class SB). You learned in Section 15-3b that the Milky Way Galaxy is a barred spiral. Massive young stars in the spiral arms give spiral galaxy disks a generally blue color. A few disk galaxies contain relatively little gas and dust and have a red tint; these are called *lenticular galaxies* (class S0), which, like ellipticals, have a red color.
3. Finally, notice the *irregular galaxies* (class I), which are generally shapeless. They tend to be rich in gas and dust and are generally blue in color, signifying ongoing star formation.

The amount of gas and dust in a galaxy strongly influences its appearance. Galaxies rich in gas and dust usually have active star formation and emission nebulae and contain hot, bright stars. That gives those galaxies a blue tint. Galaxies that are poor in gas and dust contain few or none of these highly luminous blue stars and no emission nebulae. Consequently, those galaxies have a red tint and a more uniform appearance.

Spiral galaxies are clearly disk shaped, but the true, three-dimensional shape of elliptical galaxies isn't obvious from images. Some are spherical, but the more elongated elliptical galaxies could be shaped like flattened spheres (hamburger-bun–shaped) or like elongated spheres (U.S.-football–shaped). Some may even have three different diameters; they are longer than they are thick and thicker than they are wide, like a flattened loaf of bread.

HOW DO WE KNOW? 16-1

Classification in Science

How do scientists use classification to reveal the patterns of nature?

Classification is one of the most basic and most powerful of scientific tools. Establishing a system of classification is often the first step in studying a new aspect of nature, and it can produce unexpected insights.

Charles Darwin sailed around the world from 1831 to 1836 with a scientific expedition aboard the ship HMS *Beagle*. Everywhere he went, he studied the living things he saw and tried to classify them. For example, he classified different types of finches he saw on the Galapagos Islands based on the shapes of their beaks. He found that those that fed on seeds with hard shells had thick, powerful beaks, whereas those that picked insects out of deep crevices had long, thin beaks. His classifications of these and other animals caused him to think about how natural selection shapes creatures to survive in their environment, which led him to understand how living things evolve.

Years after Darwin's work, paleontologists classified dinosaurs into two orders, lizard-hipped and bird-hipped dinosaurs. That classification, based on the shapes of dinosaur hip joints, helped scientists to understand patterns of evolution of dinosaurs. It also led to the conclusion that modern birds, including the finches that Darwin saw on the Galapagos, evolved from dinosaurs.

Astronomers use classifications of galaxies, stars, moons, and many other objects to help them see patterns, trace evolutions, and generally make sense of the astronomical world. Whenever you encounter a scientific discussion, look for the classifications on which it is based. Classifications are the orderly framework on which much of science is built.

Michael A. Seeds

The careful classification of living things has revealed that the birds, including this flamingo, are descended from dinosaurs.

Statistical studies suggest that each of these shapes is represented among elliptical galaxies.

Edwin Hubble devised the classification system for galaxies by looking at visual-wavelength images, but near-infrared images reveal the location of the majority of stars—stars that are not as luminous as the hot stars that produce most of the light. These studies suggest that the structures of many galaxies are characterized by large masses of cool, low-luminosity stars. For example, some spiral galaxies that appear to lack bars when viewed at visual wavelengths actually have massive bars that are evident in infrared images. As you think about galaxies, you should remember that their classification is based on where the light comes from and not necessarily where most of the stars are located.

It is surprisingly difficult to figure out what proportions of galaxies are elliptical versus spiral versus irregular. About 70 percent of galaxies in catalogs are spiral, but that is misleading because spiral galaxies contain hot, bright stars and clouds of ionized gas and therefore are easy to notice. Most ellipticals are fainter and harder to notice. Small galaxies such as dwarf ellipticals and dwarf irregulars are actually common but are hard to detect. (Note that this is comparable to the situation regarding stars; supergiant stars stand out and therefore seem more numerous than they actually are.) From careful statistical studies, astronomers estimate that ellipticals are actually more common than spirals, and that irregulars make up about 25 percent of all galaxies.

Why are there different kinds of galaxies? That is a key question you will explore in the rest of this chapter.

Practicing SCIENCE

Why do galaxies have different colors? To uncover the secrets of the natural world, scientists need to pay careful attention even to the simplest observations and analyze them with care.

Different kinds of galaxies have different colors, depending mostly on how much gas and dust they contain. If a galaxy contains large amounts of gas and dust, the raw material for new stars, it probably contains lots of young stars, and a few of those young stars will be massive, hot, luminous O and B stars. They will produce most of the light and give the galaxy a distinct blue tint. In contrast, a galaxy that contains little gas and dust will not have much ongoing star formation and therefore few young stars. It will lack O and B stars, and instead the most luminous stars in such a galaxy will be red giants and supergiants. They will give the galaxy a red tint.

Because the light from a galaxy is a blend of the light from billions of stars, the colors are not deep colors, only tints. Nevertheless, the most luminous stars in a galaxy determine the overall color. From this you can conclude that elliptical galaxies tend to be red, and the disks of spiral galaxies tend to be blue.

Now consider a question about galaxies that is not so simple: ***Why are most cataloged galaxies spiral despite the fact that the most common kind of galaxy is elliptical?***

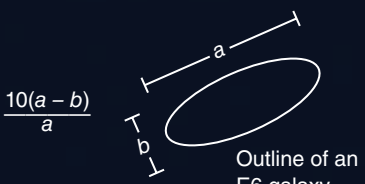
16-2 Measuring the Properties of Galaxies

Looking beyond the edge of the Milky Way Galaxy, astronomers find billions of galaxies filling space as far as anyone can see. What are the properties of these star systems—what are

Galaxy Classification

1 **Elliptical galaxies** are round or elliptical, contain no visible gas and dust, and lack hot, bright stars. They are classified with a numerical index ranging from 1 to 7; E0s are round, and E7s are highly elliptical. The index is calculated from the largest and smallest diameter of the galaxy using the following formula, rounding to the nearest integer:

$$\frac{10(a-b)}{a}$$



Outline of an E6 galaxy

NOAO/AURA/NSF

Visual

M87 is a giant elliptical galaxy classified E1. It is several times larger in diameter than our own galaxy and is surrounded by a swarm of over 500 globular clusters.

AATB/David Malin Images

Visual

The Leo 1 dwarf elliptical galaxy is not many times bigger than a globular cluster.

2 **Spiral galaxies** contain a disk and spiral arms. Their halo stars are not visible, but presumably all spiral galaxies have halos. Spirals contain gas and dust and hot, bright O and B stars, as shown at right and below. The presence of short-lived O and B stars alerts us that star formation is occurring in these galaxies. Sa galaxies have larger nuclei, less gas and dust, and fewer hot, bright stars. Sc galaxies have small nuclei, lots of gas and dust, and many hot, bright stars. Sb galaxies are intermediate. The Milky Way Galaxy is classified as Sbc, between Sb and Sc.

AATB/David Malin Images

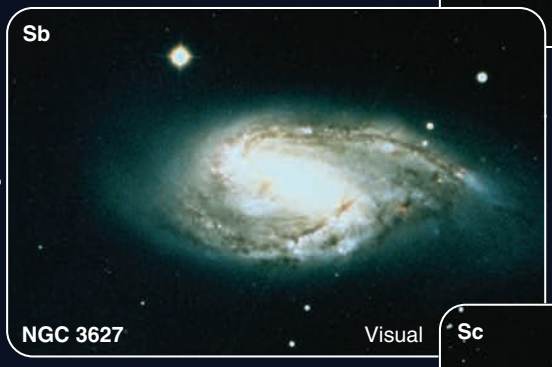
Sa

Visual NGC 3623

AATB/David Malin Images

NGC 1365 Visual

AATB/David Malin Images


Sb

NGC 3627 Visual

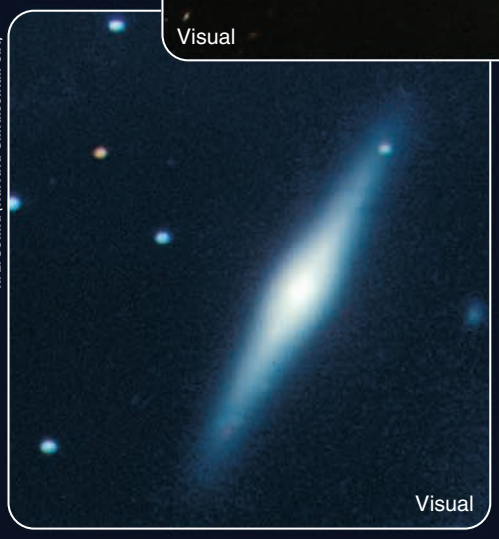
AATB/David Malin Images

Sc

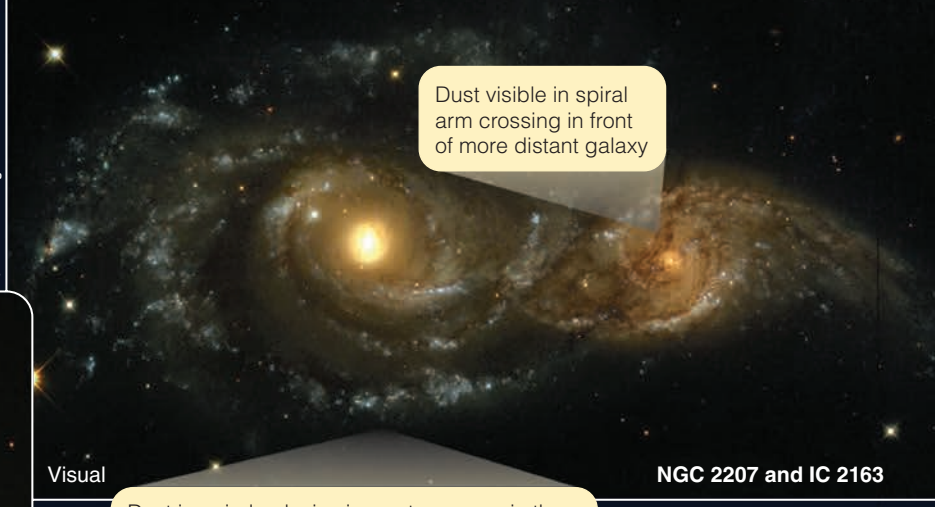
NGC 2997 Visual

2a Roughly two-thirds of all spiral galaxies are **barred spiral galaxies** classified SBa, SBb, and SBc. They have an elongated nucleus with spiral arms springing from the ends of the bar, as shown at left. The Milky Way Galaxy is a barred spiral, so its complete classification is SBbc.

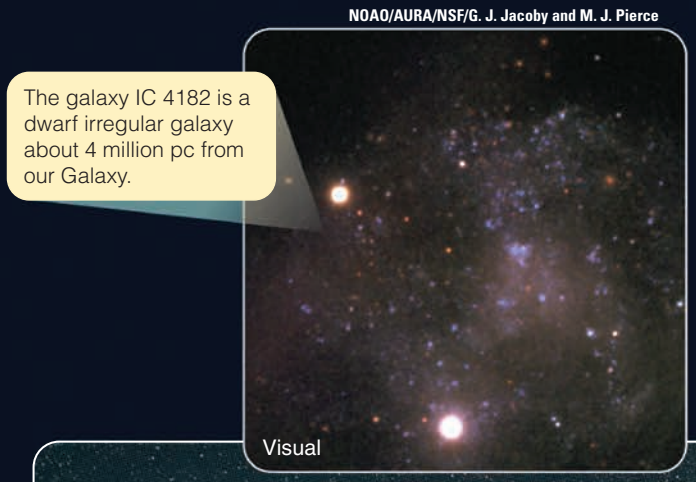
2b Some disk galaxies are rich in dust, which is concentrated along their spiral arms. NGC 4013, shown below, is a galaxy much like ours, but seen edge-on its dust is readily apparent.



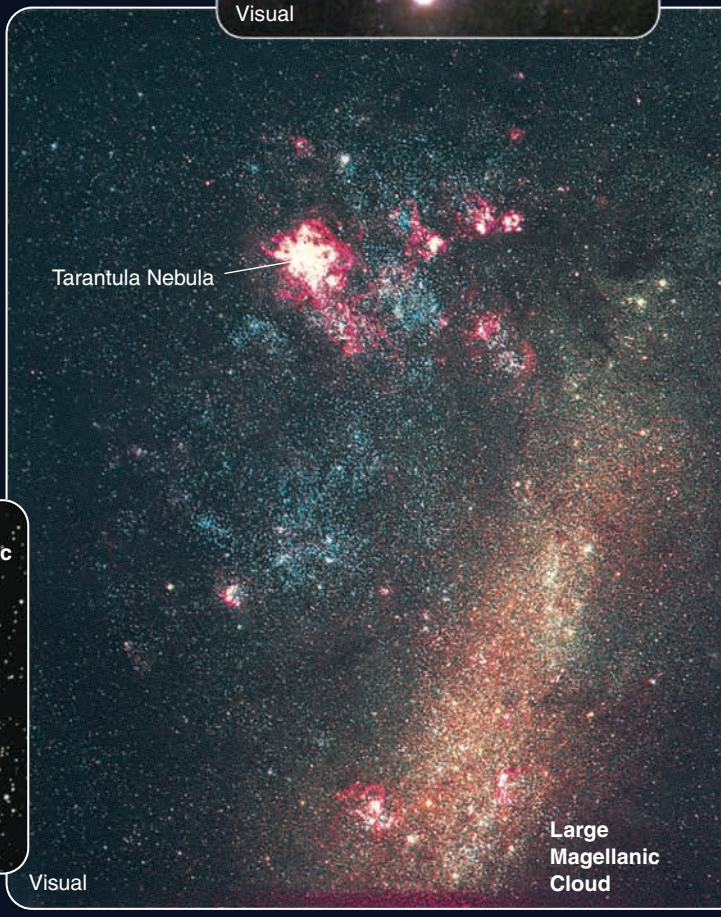
2c Galaxies with an obvious disk and nuclear bulge but little or no visible gas and dust and no hot, bright stars are called **lenticular galaxies**, classified as S0 (pronounced *Ess Zero*). Compare this galaxy with the edge-on spiral above.


NASA/ESA/STScI/AURA/NSF
/*Hubble* Heritage Team
Dust visible in spiral arm crossing in front of more distant galaxy
Visual
NGC 2207 and IC 2163

Dust in spiral galaxies is most common in the spiral arms. Here the spiral arms of one galaxy are silhouetted in front of a more distant galaxy.


NOAO/AURA/NSF/G. J. Jacoby and M. J. Pierce
The galaxy IC 4182 is a dwarf irregular galaxy about 4 million pc from our Galaxy.
Visual

3 **Irregular galaxies** (classified Irr) are a chaotic mix of gas, dust, and stars with no obvious nuclear bulge or spiral arms. The Large and Small Magellanic Clouds are visible to the unaided eye as hazy patches in the Southern Hemisphere sky. Telescopic images show that they are irregular galaxies that are interacting gravitationally with our own much larger galaxy. Star formation is rapid in the Magellanic Clouds. The bright pink regions are emission nebulae excited by newborn O and B stars. The brightest nebula in the Large Magellanic Cloud is called the Tarantula Nebula.



their diameters, luminosities, and masses? Just as in your study of stellar characteristics (Chapter 9), the necessary first step in your study of galaxies is to find out how far away they are. Once you know a galaxy's distance, its true size and luminosity are relatively easy to calculate. Later in this section, you will see that, just as for stars, finding the masses of galaxies is not easy.

16-2a Distance

The distances to galaxies are so large that it is not convenient to measure them in light-years, parsecs, or even kiloparsecs. Instead, astronomers use the unit **megaparsec (Mpc)**, or 1 million pc. One Mpc equals 3.26 million ly, or about 3.1×10^{19} km.

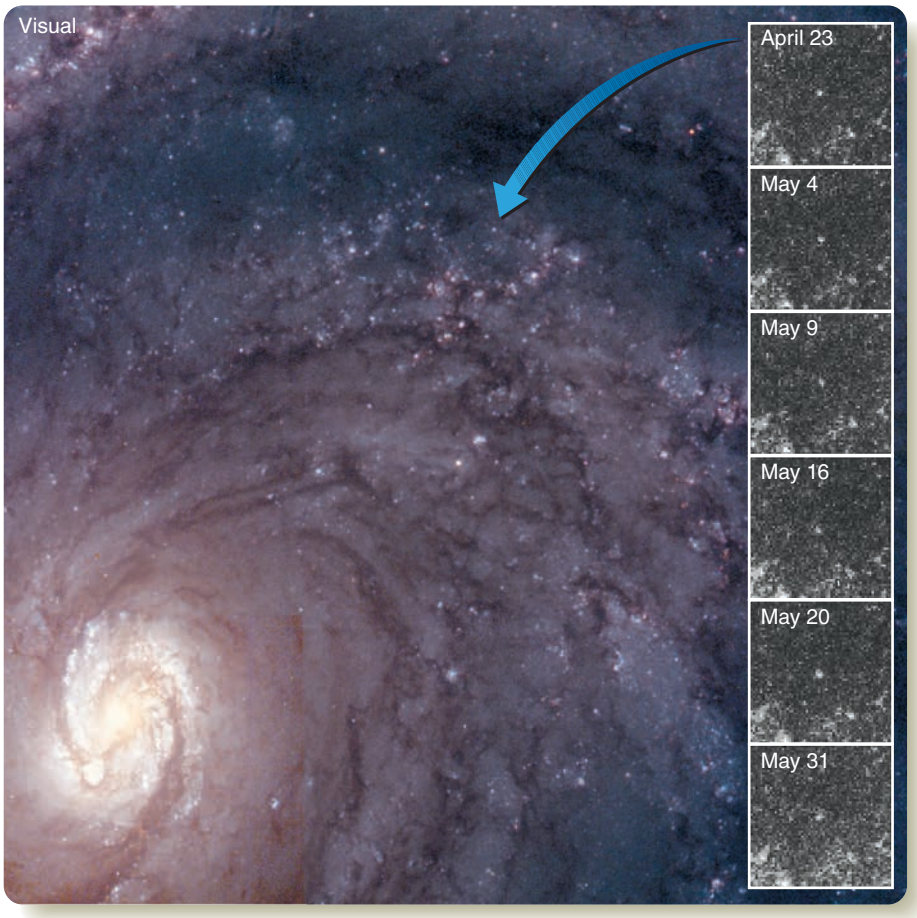
To find the distance to a galaxy, astronomers must search among its stars, nebulae, and star clusters for familiar objects with known luminosities or diameters. Such objects are called **distance indicators**; if you can find a distance indicator in a galaxy, you can estimate its distance. Astronomers sometimes refer to distance indicators with an old-fashioned term, *standard candles*.

Cepheid variable stars are reliable distance indicators because their period is related to their luminosity. The period–luminosity relation has been calibrated (look back to Figure 12-14 and also How Do We Know? 15-1), so you can use the period of the star's variation to find its absolute magnitude. Then, by comparing its absolute and apparent magnitudes, you can find its distance. Figure 16-3 shows a galaxy in which the *Hubble Space Telescope* detected Cepheids.

Even with the *Hubble Space Telescope*, Cepheids are not visible in galaxies much beyond 30 Mpc (100 million ly), so astronomers must search for less common but brighter distance indicators and calibrate them using nearby galaxies that contain both types of indicator. For example, by studying nearby galaxies with distances known from Cepheids, astronomers have found that the brightest globular clusters have absolute visual magnitudes of about –10. If you found globular clusters in a more distant galaxy, you can assume that the brightest of the globular clusters there also have absolute magnitudes of –10 and use that information to calculate the galaxy's distance.

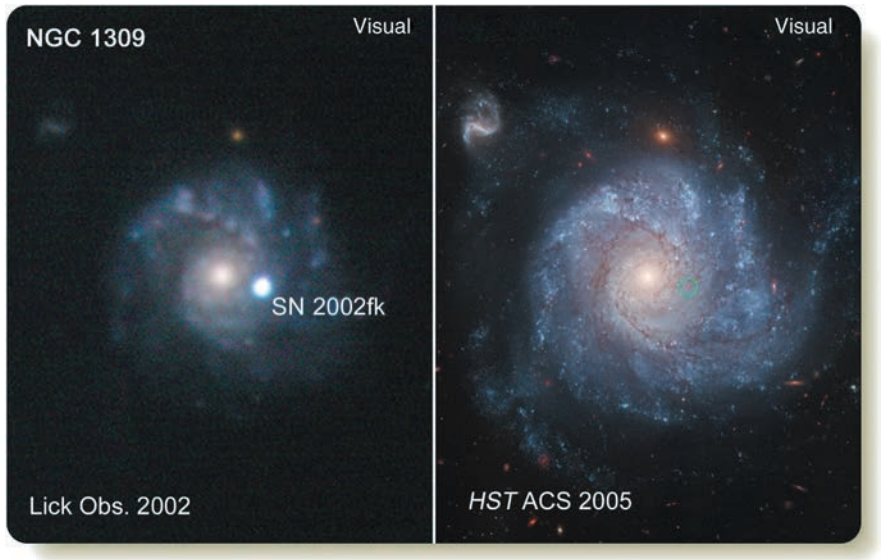
Astronomers can use globular clusters in a different way. Studies of nearby globular clusters with known distances show that they are about 25 pc in diameter. If astronomers can detect globular clusters in a distant galaxy, they can assume the clusters are about 25 pc in diameter, measure their angular diameter, and calculate the distance to the galaxy using the small-angle formula (Eq. 3-1). As you learned in Section 15-1b, an object of known size can also be a distance indicator.

Astronomers can calibrate type Ia supernovae—those produced by the collapse of a white dwarf—because the white dwarf always collapses at the same mass limit, so the explosions are similar to each other and reach approximately the same maximum luminosity. When type Ia supernovae occur in galaxies with distances that are known from Cepheid variables and other distance indicators, astronomers can find the absolute magnitude of the supernovae at peak brightness. An example is shown in Figure 16-4. When astronomers



NASA/ESA/STScI/AURA/NSF/J. Trauger (JPL-Caltech) and W. Freedman (CIW)

◀ **Figure 16-3** The vast majority of spiral galaxies are too distant for Earth-based telescopes to detect Cepheid variable stars. The *Hubble Space Telescope*, however, can locate Cepheids in some of these galaxies, as it has in the bright spiral galaxy M100. From a series of images taken on different dates, astronomers can locate Cepheids (*inset*), determine the period of pulsation, and measure the average apparent brightness. They can then deduce the distance to the galaxy: 17 Mpc (55 million ly) for M100.



Left panel: UCO/W. Li and A. Filippenko (Univ. of California, Berkeley); Right panel: NASA/ESA/STScI/AURA/NSF/A. Riess

▲ **Figure 16-4** While dinosaurs roamed Earth, a white dwarf in the galaxy NGC 1309 collapsed and exploded as a type Ia supernova, and the light from that explosion reached Earth in 2002. Astronomers found Cepheid variable stars in the galaxy, so they can determine that it is 100 million ly away, and that allowed them to find the absolute magnitude of the supernova at its brightest. The supernova had faded from view three years later, even though the second image was made using the *Hubble Space Telescope*. By combining observations of many supernovae, astronomers have been able to calibrate type Ia supernovae as distance indicators.

see a type Ia supernova in a galaxy too far away to use other distance indicators, they measure the apparent magnitude of the supernova at maximum, compare that with the known absolute magnitude these supernovae reach at maximum, and find the distance to the galaxy. As you will see in a later chapter, this is a critical calibration involved in modern measurements of the size and age of the Universe.

Cepheids and globular clusters are invisible in the most distant galaxies, and supernovae are rare. Consequently, to measure the greatest distances, astronomers have calibrated the total luminosity of the galaxies themselves. For example, studies of nearby galaxies show that a normal large spiral galaxy like the Milky Way has a luminosity of about 16 billion times that of the Sun. If you see a spiral galaxy far away, you can measure its apparent magnitude and estimate its distance. Of course, it is important to recognize the different types of galaxies, and that is difficult to do at great distances, as you can see by looking at the fainter galaxies in Figure 16-2. Averaging the estimated distances to several of the brightest spiral galaxies in a cluster can reduce the uncertainty in this method.

Notice how astronomers use calibration to build a **distance scale**, reaching from the nearest galaxies to the most distant visible galaxies. Sometimes astronomers refer to this as the "distance ladder" because each step depends on the steps below it. The most dependable step in measuring galaxy distances is the Cepheid variable stars, but notice that Cepheid distance indicators depend on astronomers' understanding of the luminosities of the stars in the H–R diagram, and that rests on a lower step in the ladder—measurements of the parallax of stars. Stellar parallaxes in turn depend on measuring the size of Earth's orbit around the Sun, which finally rests on the bottom step, measuring the size of Earth itself. It's worth a moment's pause to consider how the distance ladder ultimately connects the size of Earth to the most distant galaxies in the Universe.

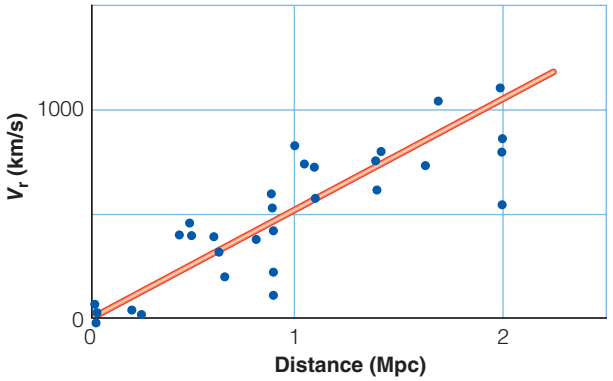
The most distant known galaxies are more than 3000 Mpc (10 billion ly) away, and at such distances you see an effect resembling time travel. When you look at a galaxy that is millions of light-years away, you do not see it as it is now but as it was millions of years ago when its light began the journey toward Earth. When you look at a distant galaxy, you look back into the past by an amount called the **look-back time**, which is a time in years equal to the distance in light-years the light from the galaxy traveled to reach Earth.

The look-back time to nearby objects is usually not significant. For example, the look-back time across a football field is less than one-millionth of a second. The look-back time to the Moon is only 1.3 seconds, to the Sun 8 minutes, and to the nearest star a bit over 4 years. The Andromeda Galaxy has a look-back time of more than 2 million years. That seems like a long time in human terms, but it is a mere eyeblink in the lifetime of a galaxy, so you can be sure that "right now" the Andromeda Galaxy is pretty much as it appears to us. However, if you look at more distant galaxies, the look-back time becomes an appreciable fraction of the age of the Universe. You will learn in a later chapter about evidence that the Universe began almost 14 billion years ago. When you look at the most distant visible galaxies, you are looking back over 10 billion years to a time when the Universe was significantly different. The look-back time becomes an important factor as you begin to think about the origin and evolution of galaxies.

16-2b The Hubble Law

Although astronomers must work carefully to measure the distance to a galaxy, they often estimate such distances using a simple relationship that was first noticed at about the same time astronomers were beginning to understand the nature of galaxies.

In 1913, Vesto Slipher at Lowell Observatory reported on the spectra of some of the faint spiral nebulae. Their spectra seemed to be composed of a mixture of stellar spectra. Some had Doppler shifts that suggested rotation; most had redshifts as if they were receding; and the faintest had the largest redshifts. Within two decades, astronomers concluded that the faint objects are galaxies similar to our own Milky Way and that the galaxies are indeed receding from us in a general expansion.



▲ **Figure 16-5** Edwin Hubble's first diagram of the apparent velocities of recession and distances of galaxies did not probe very deeply into space, and the horizontal axis scale was later recalibrated. These data did show, however, that the galaxies are receding from one another.

In the 1920s, astronomers Edwin Hubble and Milton Humason were able to measure the distances to a number of galaxies using Cepheid variable stars. In 1929, they published a graph that plotted the apparent velocity of recession versus distance for their galaxies. The points in the graph fell along a straight line (**Figure 16-5**). The straight-line relation from Hubble's diagram can be written as a simple equation:

KEY EQUATION 16-1 The Hubble Law

$$V = Hd$$

That is, the apparent velocity of recession V equals the distance d in millions of parsecs times the constant H. This relation between redshift and distance is known as the **Hubble law**, and the constant H is known as the **Hubble constant**. Modern measurements of distance show that H equals about 70 km/s/Mpc. (H has units of velocity divided by distance. These are usually written as km/s/Mpc, meaning kilometers per second per megaparsec.)

Simply stated, a galaxy's distance equals its apparent velocity of recession divided by H. For example, if a galaxy has an apparent radial velocity of 700 km/s, and H is 70 km/s/Mpc, then the distance to the galaxy is 700 divided by 70, or about 10 Mpc. This makes it relatively easy to estimate the distances to galaxies because a large telescope can photograph the spectrum of a galaxy and determine its apparent velocity of recession even when it is too distant to have detectable distance indicators.

The Hubble law has important implications. It is interpreted to mean that the Universe is expanding. You will study this expansion in a later chapter, but here you can use the Hubble law first as a practical way to estimate the distance to a galaxy.

16-2c Diameter and Luminosity

Once you find the distance to a galaxy from distance indicators or the Hubble law, you can calculate its diameter and its luminosity. With a good telescope and the right equipment, you could easily photograph a galaxy and measure its angular diameter in arc seconds. If you knew the distance, you could use the small-angle formula (Eq. 3-1) to find its linear diameter. If you also measured the apparent magnitude of a galaxy, you could use the magnitude–distance formula (Eq. 9-2a) to find its absolute magnitude and from that its luminosity.

The results of such observations show that galaxies differ dramatically in size and luminosity. Irregular galaxies tend to be small—5 to 25 percent the size of our Galaxy—and of low luminosity. Although they are common, they are easy to overlook. Our Milky Way Galaxy is large and luminous compared with most spiral galaxies, though astronomers know of a few spiral galaxies that are even larger and more luminous. The largest is nearly five times bigger in diameter and about ten times more luminous than the Milky Way. Elliptical galaxies cover a wide range of diameters and luminosities. The largest, called giant ellipticals, are about ten times the diameter of our Milky Way, but many elliptical galaxies are small dwarf ellipticals that are only 1 percent the diameter of our Galaxy.

You can use an analogy to get a better sense of proportion. If our Galaxy were a minivan, the smallest dwarf galaxies would be the size of pocket-size toy cars, and the largest giant ellipticals would be the size of passenger jets. Even among spiral galaxies that appear similar in shape, the sizes can range from that of a small bicycle to that of a large bus (**Figure 16-6**).

Clearly, the diameter and luminosity of a galaxy do not determine its type. Some small galaxies are irregular, and some

▲ **Figure 16-6** The small galaxy cluster known as Hickson Compact Group 87 appears to contain three spiral galaxies and one elliptical. It is not known whether the small spiral galaxy is in the distant background or is part of the group. Note the contrast between the sizes of the galaxies.

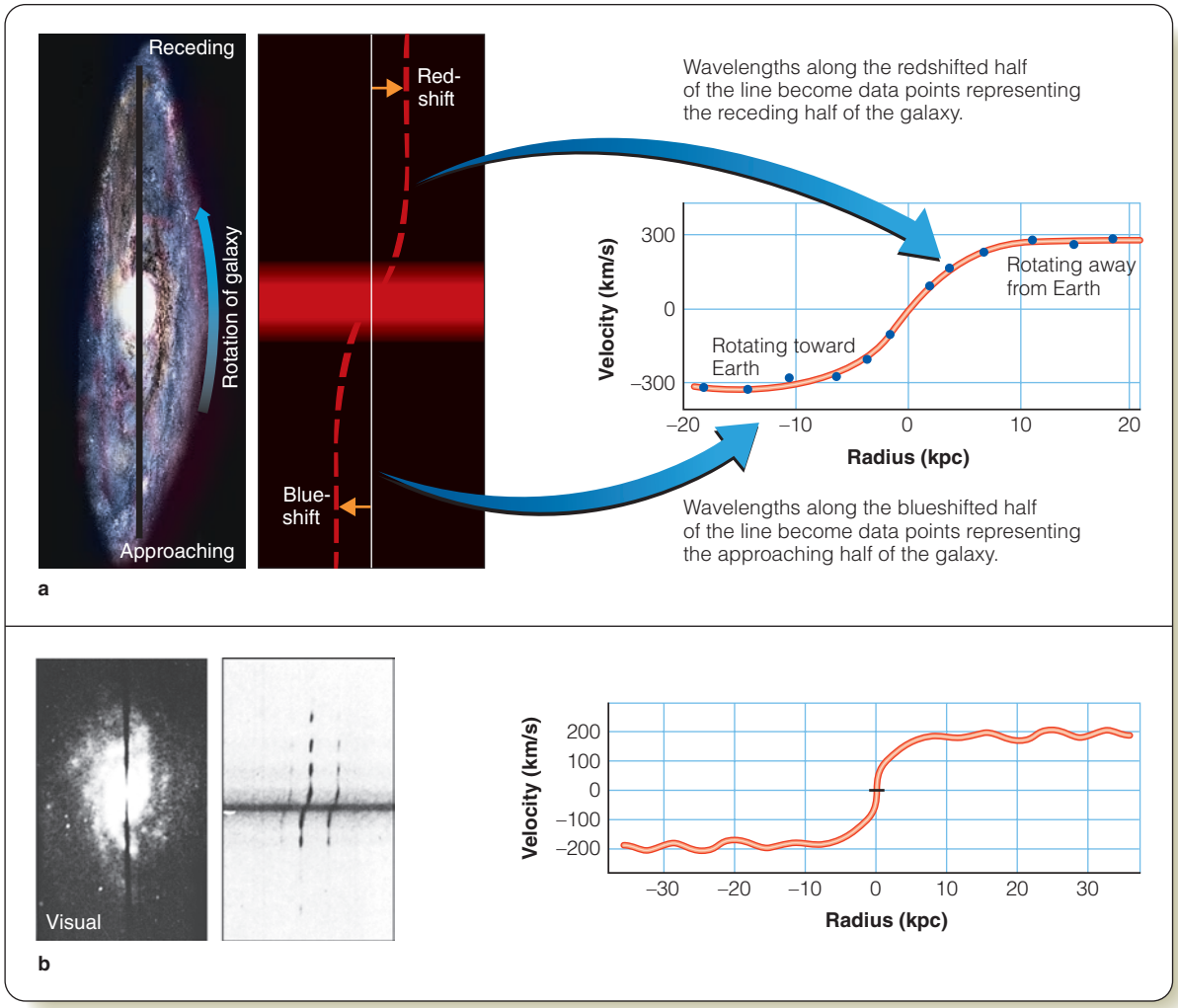
are elliptical. Some large galaxies are spiral, and some are elliptical. Later in this chapter, you can build a hypothesis for the origin and evolution of galaxies, but first you need to add a third basic parameter—mass.

16-2d Mass

Although the mass of a galaxy is difficult to determine, it is an important quantity. It tells you how much matter the galaxy contains, which in turn provides clues to the galaxy's origin and evolution. In this section, you will examine two fundamental ways to find the masses of galaxies.

One way to find the mass of a galaxy is to watch it rotate. All you need to know is the radius of the galaxy and the orbital period of the stars orbiting at the galaxy's outer edge. Then you can use Kepler's third law to find the mass, just as you found the mass of binary stars in Section 9-5a.

Of course, you don't live long enough to see a galaxy rotate, so you must find the stars' orbital periods using a chain of inference. You could find the orbital velocity of the stars from the Doppler effect. If you focus the image of the galaxy on the slit of a spectrograph, you see a bright spectrum formed by the bright nucleus of the galaxy, in addition to fainter emission lines produced by ionized gas in the disk of the galaxy. Because the galaxy rotates, one side moves away from Earth and one side moves toward Earth, so the emission lines would be redshifted on one side of the galaxy and blueshifted on the other side. You could measure those changes in wavelength, use the Doppler formula to find the velocities, and plot a diagram showing the velocity of rotation at different distances from the center of the galaxy—a diagram called a *rotation curve* (Section 15-2b). The artwork in **Figure 16-7a** shows the process of creating a rotation curve, and **Figure 16-7b** shows a real galaxy, its spectrum, and its rotation curve.



▲ **Figure 16-7** (a) An astronomer has placed the image of the galaxy over a narrow slit so that light from the galaxy can enter the spectrograph and produce a spectrum. A very short segment of the spectrum shows an emission line redshifted on the receding side of the rotating galaxy and blueshifted on the approaching side. Converting those Doppler shifts into velocities, the astronomer can plot the galaxy's rotation curve (*right*). (b) Galaxy NGC 2998 is shown over the spectrograph slit, and the segment of the spectrum includes three emission lines.

Of course, once you know the orbital velocity and the radius of a star's orbit around a galaxy, you can calculate its orbital period. You followed this procedure in Section 15-2b to estimate the mass of the Milky Way Galaxy. The circumference of the orbit (2π times the radius) divided by the orbital velocity equals the orbital period. So knowing the velocity and size of an orbit tells you the orbital period, and from that you can find the mass. That is why the rotation curve of a galaxy is so important; it contains all the information you need to find the mass of the galaxy, and the method is known as the **rotation curve method**. It is the most accurate way to find the mass, but it works only for nearby galaxies whose rotation curves can be measured easily. More distant galaxies appear so small that astronomers cannot measure the radial velocity at different points across the galaxy.

You should note one warning about masses found from rotation curves. Studies of our own Galaxy and other galaxies show that the outer parts of their rotation curves do not decline to lower velocities as you would expect if most of the mass lies in the inner parts of each galaxy. As in the case of the rotation curve of the Milky Way Galaxy (look back to Figure 15-10), this indicates that the galaxies contain large amounts of mass in their outer portions, in extended galactic coronae. Because the rotation curve method can be applied only to nearby galaxies and because it cannot determine the masses of galactic coronae, astronomers have looked at another way to find the total masses of galaxies.

A related way of measuring a galaxy's mass is called the **velocity dispersion method**. In the spectra of some galaxies, broad spectral lines indicate that stars and gas are moving at high velocities. If you assume the galaxy is bound by its own gravity and is not coming apart, you can ask how massive it must be so that its escape velocity is larger than the velocity of the fastest object in the galaxy.

The **cluster method** of finding the total mass of a cluster of galaxies depends on measuring the motions of galaxies within the cluster. It is a version of the velocity dispersion method described in the previous paragraph, applied to a group of galaxies. If you measured the radial velocities of many individual galaxies in a cluster, you would find that some velocities are larger than others. You could ask how massive a cluster of that size must be to hold itself together, given the range of velocities in the cluster. That is the same as asking how massive the cluster must be to have an escape velocity larger than the velocity of the fastest galaxy. Solving this problem would tell you the mass of the entire cluster. Dividing the total mass of the cluster by the number of galaxies in the cluster yields the average mass of the galaxies. Note that this method contains the built-in assumption that the cluster is stable. If it is actually coming apart, your calculated result is too large. Because most galaxies are found in clusters it seems likely that the clusters are stable structures, held together by their own gravity. Therefore, the cluster method is probably valid.

Table 16-1 Properties of Galaxies*

	Elliptical	Spiral	Irregular
Mass	0.0001–50	0.005–5	0.0005–0.15
Diameter	0.01–10	0.2–5	0.05–0.25
Luminosity	0.00005–5	0.005–10	0.00005–0.1

*In units of the mass, diameter, and luminosity of the Milky Way Galaxy.

Measuring the masses of galaxies reveals two things. First, the range of masses is wide—from 10,000 times smaller than the Milky Way Galaxy to 50 times larger (Table 16-1). Second, galaxies contain invisible dark matter spread through extended galactic coronae, just as does our Milky Way Galaxy. You will learn more about dark matter later in this chapter.

One more important aspect of galaxies remains to be explored. What lies in their centers? The Milky Way Galaxy contains a 4-million-solar mass ($M_\odot$) black hole at its center. Do other galaxies contain similar objects?

16-2e Supermassive Black Holes in Galaxies

Rotation curves show the motions of the outer parts of a galaxy, but it is also possible to detect the Doppler shifts of stars orbiting close to the centers. Although these motions are not usually shown on rotation curves, they reveal something astonishing.

Measurements show that the stars near the centers of many galaxies are orbiting surprisingly rapidly. To hold stars in such small, short-period orbits, the centers of galaxies must contain millions or even billions of solar masses in a small region. The evidence shows that the nuclei of many galaxies contain supermassive black holes. The Milky Way contains a supermassive black hole at its center (Section 15-4), and evidently that is typical of galaxies.

A black hole with such an enormous mass cannot be the remains of a dead star. Stellar black holes with measured masses are all smaller than 20 $M_\odot$. Measurements show that the mass of a supermassive black hole in the center of a galaxy is proportional to the mass of its host galaxy's central bulge. Galaxies with large central bulges tend to have massive central black holes, and galaxies with small central bulges have relatively small central black holes. Rare galaxies without central bulges usually also lack supermassive black holes (Figure 16-8). This suggests that the supermassive black holes formed long ago as part of the process of the galaxies and their central bulges forming. Matter has probably continued to drain into the central black holes from their surrounding galaxies, but the consistent relationship between bulge mass and black hole mass indicates the black holes have not grown dramatically since they formed.

A 1-billion-$M_\odot$ black hole sounds like a lot of mass, but note that it is less than 1 percent of the mass of a galaxy as large as the

AURA/NOAO/NSF/W. Schoening

◀ **Figure 16-8** The galaxy M33 is near our Milky Way Galaxy, and it can be studied in detail. The velocities at the center of the galaxy are low, showing that it does not contain a supermassive black hole at its center. It also lacks a central bulge, and that confirms the observation that, with few exceptions, the mass of a galaxy's central supermassive black hole is related to the size of that galaxy's central bulge.

Milky Way. The 4-million-$M_\odot$ black hole at the center of the Milky Way Galaxy contains only one-thousandth of 1 percent of the mass of the entire Galaxy. In the next chapter, you will discover that although these supermassive black holes have small masses compared with the masses of their host galaxies, they can nevertheless produce titanic eruptions.

16-2f Dark Matter in Galaxies

Given the luminosity of a galaxy, astronomers can make a rough guess about the amount of matter it should contain. From visible and near-infrared observations they know how much light all the stars produce, and from far-infrared observations they know about how much matter there is in the interstellar medium, so it should be possible to estimate the total mass of a galaxy from the radiation it emits. But when astronomers compare those estimates with masses of galaxies estimated from the effects of their gravitation, they find that the measured masses are generally about ten times larger than the masses that can be seen. This must mean that nearly all galaxies contain dark matter. The dark matter must be made up of some as-yet-undiscovered type or types of particles that do not interact with light, normal matter, or with each other, and are detectable only through their gravitational fields.

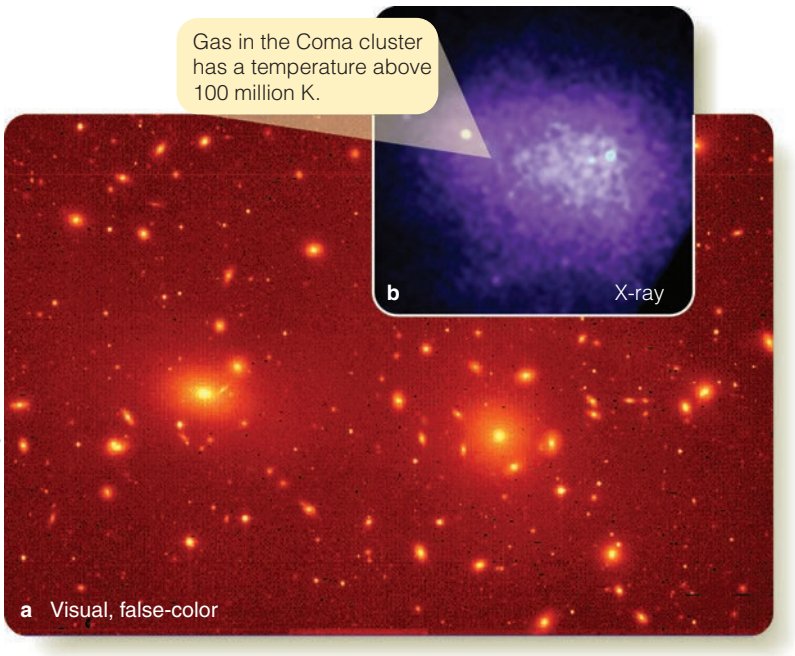
The effects of dark matter are difficult to detect, and even harder to explain. The halos of galaxies contain faint objects such as white dwarfs, red dwarfs, and brown dwarfs, but not nearly enough to be the dark matter. It can't be black holes and neutron stars because the X-rays those objects would emit are not detected. There must be roughly ten times more dark matter than visible matter, and such a huge number of black holes and neutron stars would produce X-rays that would be easy to detect.

Ironically, X-ray observations do reveal some evidence of dark matter. X-ray images of galaxy clusters show that many of them are filled with hot, low-density gas. The amount of gas present is much too small to account for the dark matter. Rather, the gas is important because it is very hot but its rapidly moving atoms have not escaped. Evidently the gas is held in the cluster by a strong gravitational field. To have a high enough escape velocity to hold the hot gas, the cluster must contain much more matter than what astronomers can observe directly. The detectable galaxies in the Coma cluster, for instance, amount to only a small fraction of the total mass of the cluster (**Figure 16-9**).

Dark matter is surprisingly important in the composition of the Universe. Observations of galaxies and clusters of galaxies show that approximately 90 percent of the matter in the Universe must be dark matter. The Universe you see—the kind of matter that you and the stars are made of—has been compared to the foam on an invisible ocean.

16-2g Gravitational Lensing and Dark Matter

When you solve a math problem and get an answer that doesn't seem right, you check your work for a mistake. Astronomers



Panel a: G. Bothun (Univ. of Oregon); Panel b: NASA/CXC/SAO/A. Vikhlinin et al.

▲ **Figure 16-9** (a) The Coma cluster of galaxies contains at least 1000 galaxies and is especially rich in E and S0 galaxies. Two giant galaxies lie near its center. Only the central area of the cluster is shown in this image. If the cluster were visible in the sky, it would span eight times the diameter of the full moon. (b) This X-ray image of the Coma cluster shows that it is filled and surrounded by hot gas. Note that the two brightest galaxies are visible in the X-ray image.

using Newton's laws (for example, calculations of orbital velocity and escape velocity) to measure the mass of galaxies find evidence of large amounts of mysterious dark matter, so they look for ways to check that result. Dark matter is so peculiar that a few scientists have speculated it isn't real and that some quirk of gravity makes it just seem to be there. Fortunately, there is another way to detect dark matter that does not depend on Newton's laws. Astronomers can study the tracks of light beams.

Albert Einstein described gravity as a curvature of space-time that is caused by the presence of mass (look back to Section 5-3c). Einstein predicted that a light beam traveling through a gravitational field would be deflected by the curvature of space-time much as a golf ball is deflected as it rolls over a curved putting green. That effect has been observed to have the size predicted by Einstein, a strong confirmation that Einstein's theory is correct.

Gravitational lensing occurs when light from a distant object passes a nearby massive object and is deflected by the gravitational field. The gravitational field of the nearby object is actually a region of curved space-time that acts as a lens to deflect the passing light. Astronomers can use gravitational lensing to detect dark matter when light from very distant galaxies passes through a closer cluster of galaxies on its way to Earth and is deflected by the strong curvature. The deflection can produce multiple images of the distant galaxies and distort them into arcs. The amount of the distortion depends on the mass of the nearer cluster of galaxies (**Figure 16-10a**). Observations of gravitational lensing made with very large telescopes reveal that clusters of galaxies contain far more matter than what can be seen. That is, they contain large amounts of dark matter. This confirmation of the existence of dark matter is independent of Newton's laws and gives astronomers great confidence that dark matter is real.

In another example, astronomers have identified two clusters of galaxies that passed through each other about 100 million years ago. X-ray observations show that the hot gas in the clusters collided and was swept out of the clusters by the collision, but gravitational lensing measurements show that the clusters still contain dark matter. As predicted by theory, the dark matter in the two clusters did not collide (did not interact) as the gas in the two clusters did, and so the dark matter remained with the clusters that are now moving in opposite directions from the point of collision. You can see the gas as the pink areas in **Figure 16-10b** and the locations of the dark matter as the purple areas. This is compelling evidence that dark matter really exists and is not an unexpected, subtle feature of gravity.

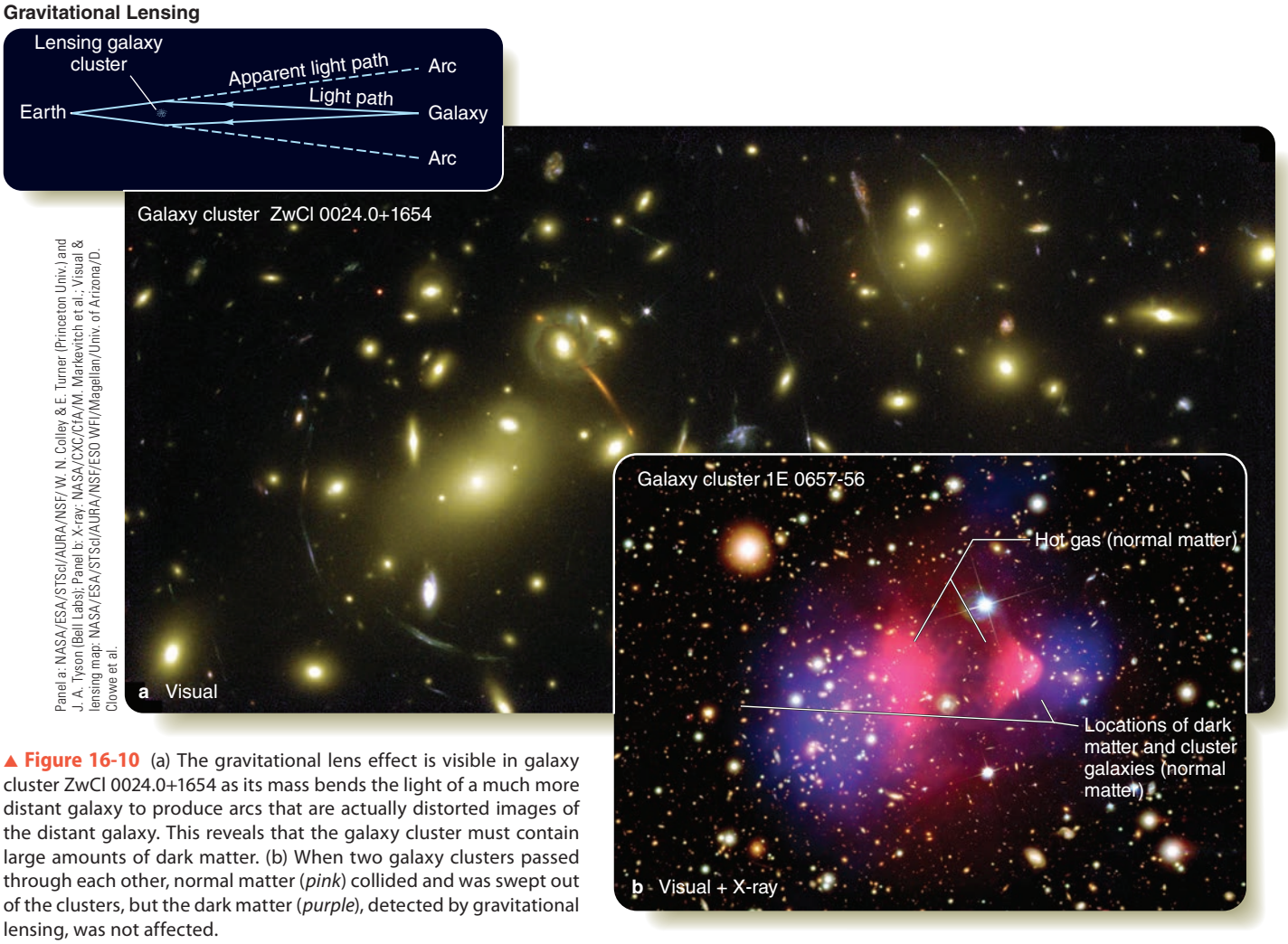
Dark matter remains one of the fundamental unresolved problems of modern astronomy. Some physicists are attempting to detect dark matter particles in the vicinity of Earth with an instrument on the *International Space Station* or in ground-based laboratory experiments, so far without success. You will learn more about this problem in a later chapter when you try to understand how dark matter affects the nature of the Universe, its past, and its future.

Practicing **SCIENCE**

Why do you have to know the distance to a galaxy to find its mass? Here is a good example of a chain of inference needed to determine a desired bit of information from measurements of other quantities.

To find the mass of a galaxy, you need to know the size of the orbits and the orbital periods of stars at the galaxy's outer edge and then use Kepler's third law to find the mass inside the orbits. Measuring the orbital velocity of the stars as the galaxy rotates is easy if you can obtain a rotation curve from a spectrum. But you must also know the radii of the stars' orbits in linear units such as meters or AU. That is where the distance comes in. Once you know the distance to the galaxy, you can use the small-angle formula to convert the radius in arc seconds into a radius in pc, in AU, or in meters. If your measurement of the distance to the galaxy isn't accurate, you will get inaccurate radii for the orbits and will compute an inaccurate mass for the galaxy.

Many different measurements in astronomy depend on the calibration of the distance scale. ***What would happen to your measurements of the diameters and luminosities of the galaxies if astronomers discovered that the Cepheid variable stars are slightly more luminous than had been thought?***



Panel a: NASA/ESA/STScI/AURA/NSF/ W. N. Colley & E. Turner (Princeton Univ.) and J. A. Tyson (Bell Labs); Panel b: X-ray: NASA/CXC/CfA/M. Markevitch et al.; Visual & lensing map: NASA/ESA/STScI/AURA/NSF/ESO WFI/Magellan/Univ. of Arizona/D. Clowe et al.

▲ **Figure 16-10** (a) The gravitational lens effect is visible in galaxy cluster ZwCl 0024.0+1654 as its mass bends the light of a much more distant galaxy to produce arcs that are actually distorted images of the distant galaxy. This reveals that the galaxy cluster must contain large amounts of dark matter. (b) When two galaxy clusters passed through each other, normal matter (*pink*) collided and was swept out of the clusters, but the dark matter (*purple*), detected by gravitational lensing, was not affected.

16-3 Evolution of Galaxies

Your goal in this section is to use observations and theory to explain the evolution of galaxies. In Section 15-5, you considered the origin of our Milky Way Galaxy; presumably, other galaxies formed similarly. But why did some galaxies become spiral, some elliptical, and some irregular? An important clue to that mystery lies in the clustering of galaxies.

16-3a Clusters of Galaxies

Single, isolated galaxies are rare. Instead, most galaxies occur in clusters containing a few to a few thousand galaxies spread across volumes 1 to 10 Mpc in diameter (**Figure 16-11**). Surveys have cataloged thousands of galaxy clusters.

For purposes of this study, you can sort clusters of galaxies into rich clusters and poor clusters. **Rich clusters** contain a thousand or more galaxies, mostly ellipticals, scattered through a volume roughly 3 Mpc (10^7 ly) in diameter. Such a cluster is nearly always condensed; that is, the galaxies are more crowded near the cluster center. At their centers, such clusters often contain one or more giant elliptical galaxies.

The Virgo cluster is an example of a rich cluster. It contains more than 2500 galaxies and is about 17 Mpc (54 million ly) from the Milky Way. The Virgo cluster is centrally condensed and contains the giant elliptical galaxy M87 at its center. Many of these rich clusters are filled with a hot gas—an intracluster medium—that is revealed by X-ray observations. Some of this gas has presumably been driven out of the cluster galaxies by supernovae explosions, but much of it appears to be left over from the formation of the cluster.

Poor clusters contain a few dozen to a few hundred galaxies spread relatively sparsely through a region that can be as large as a rich cluster. The Milky Way Galaxy is a member of a poor cluster known by the unimaginative name of the **Local Group** (**Figure 16-12a**) containing about 40 galaxies scattered irregularly through a volume roughly 1 Mpc in diameter. Of the brighter galaxies, 15 are elliptical, 4 are spiral, and 13 are irregular.

NOAO/AURA/NSF

◀ **Figure 16-11** The Hercules galaxy cluster is named after the constellation in which it is found. It is a small cluster containing roughly a hundred galaxies, both elliptical and spiral. It lies more than 500 million ly from our Galaxy.

The total number of galaxies in the Local Group is uncertain because some galaxies lie in the plane of the Milky Way Galaxy and are difficult to detect. For example, the Sagittarius Dwarf Galaxy lies on the far side of our own Galaxy, where it is almost totally hidden behind the star clouds of Sagittarius (**Figure 16-12b**). Even closer to the Milky Way Galaxy is the Canis Major Dwarf Galaxy (**Figure 16-12c**). This galaxy was found by mapping the distribution of red supergiants detected by the 2MASS infrared all-sky survey. There are certainly other small galaxies in our Local Group that have not been detected yet.

Poor galaxy clusters tend not to be centrally condensed; rather, they tend to have subclusters with smaller galaxies crowded around larger galaxies. The Local Group illustrates this subclustering. The two largest galaxies, the Milky Way Galaxy and the Andromeda Galaxy, are the centers of two subclusters. The Milky Way Galaxy is accompanied by the Magellanic Clouds and at least seven other dwarf galaxies. The Andromeda Galaxy is attended by more dwarf elliptical galaxies and a small spiral galaxy, M33 (Figure 16-12a). The Andromeda Galaxy and our Milky Way Galaxy, with their retinues of smaller galaxies, are known to be moving toward each other and will almost certainly collide within a few billion years.

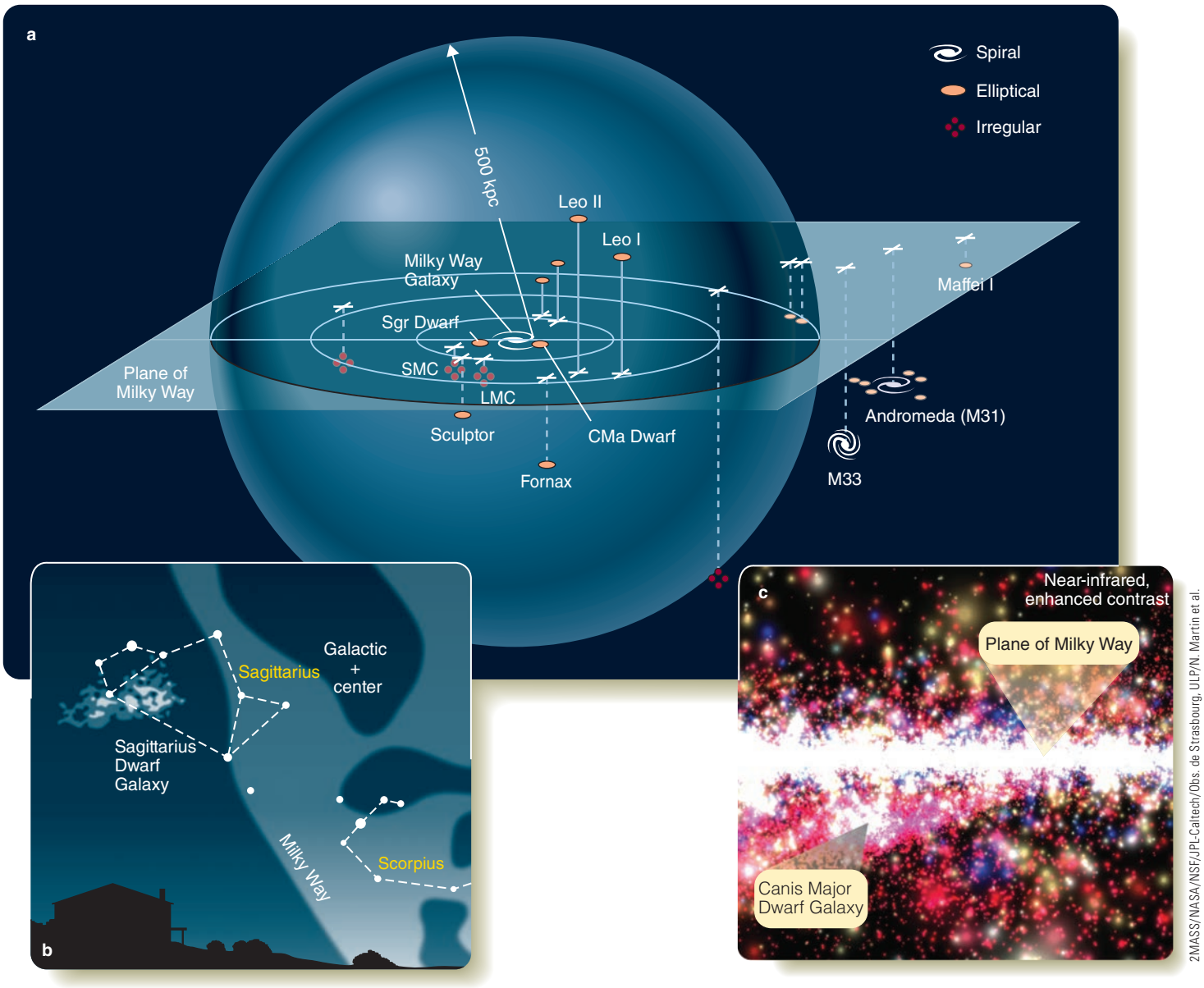
From all of this, you can draw an important conclusion: Galaxies do not live in isolation. They are usually found in clusters, often interact with nearby companions, and may even collide with each other. That leads to an important clue about galaxy evolution that is hidden in the clusters of galaxies—the relative abundance of the different types of galaxies.

In general, rich clusters normally contain 80 to 90 percent E and S0 galaxies and a few spirals. In other words, galaxies that are crowded together tend to be E or S0 rather than spirals. Poor clusters contain a larger percentage of spirals. Among the rare isolated galaxies that are not in clusters, 80 to 90 percent are spirals. Somehow the environment around a galaxy helps determine its type.

Astronomers suspect that collisions between galaxies are an important process. In a rich cluster, the galaxies are much closer together, and they must collide with each other more often than galaxies in a poor cluster. Could such collisions explain the excess of elliptical galaxies in rich clusters? In fact, astronomers have discovered that galaxy smashups can radically change the shapes of the galaxies involved.

16-3b Colliding Galaxies

Stars almost never collide. In the region of the Milky Way near the Sun, the average separation between stars is about 10^7 times their diameters. Consequently, a collision between two stars inside a galaxy is about as likely as collision between two gnats flitting about in a football stadium. On the other hand, galaxies should collide fairly often. The average separation between galaxies is only about 2 to 20 times their diameter, depending on whether they are in a rich or poor cluster. Like two blindfolded elephants



▲ **Figure 16-12** (a) The Local Group. Our Galaxy is located at the center of this diagram. The vertical lines giving distances from the plane of the Milky Way are solid above the plane and dashed below. (b) The Sagittarius Dwarf Galaxy (Sgr Dwarf) lies on the other side of our Galaxy. If you could see it in the sky, its minor axis would appear 17 times wider than the full Moon. (c) The Canis Major Dwarf Galaxy (CMa Dwarf), shown in this simulation, is even closer to the Milky Way Galaxy. It is the remains of a small galaxy that is being pulled apart by our Galaxy and is hidden behind the stars of the constellation Canis Major.

blundering about under a circus tent, galaxies should bump into each other fairly often.

Study **Interacting Galaxies** and notice four important points and three new terms:

1. Interacting galaxies can distort each other with tides that produce *tidal tails* and shells of stars. The tides may even trigger the formation of spiral arms. Large galaxies can even absorb smaller galaxies, a process called *galactic cannibalism*.
2. Interactions between galaxies can trigger rapid star formation.
3. Evidence inside some galaxies in the form of conflicting directions of star and gas motions and multiple nuclei reveals that they have suffered interactions in the past that led to mergers.
4. The beautiful *ring galaxies* are understood to be bull's-eyes left behind by high-speed face-on collisions.

Evidence of galaxy mergers is all around. Our Milky Way Galaxy is actually a cannibal galaxy snacking on the nearby Magellanic Clouds. Furthermore, our Galaxy's tides are pulling the Sagittarius Dwarf Galaxy apart, and the Canis Major Dwarf Galaxy has been almost completely digested as tides have pulled stars away to form great streamers wrapped around the Milky Way (**Figure 16-13**). Our Galaxy has almost certainly dined on other small galaxies in the past.

Interacting Galaxies

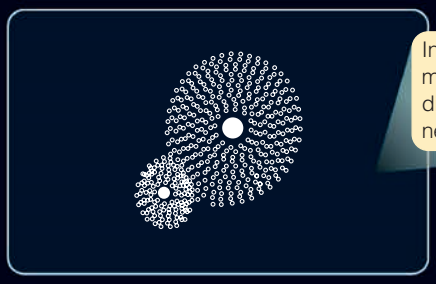
1 When two galaxies collide, they can pass through each other without stars colliding because the stars are so far apart relative to their sizes. Gas clouds and magnetic fields do collide, but the biggest effects may be tidal. Even when two galaxies just pass near each other, tides can cause dramatic effects, such as long streamers called **tidal tails**. In some cases, two galaxies can merge and form a single galaxy.

Tidal Distortion
Small galaxy passing near a massive galaxy.
Gravity of a second galaxy represented as a single massive object

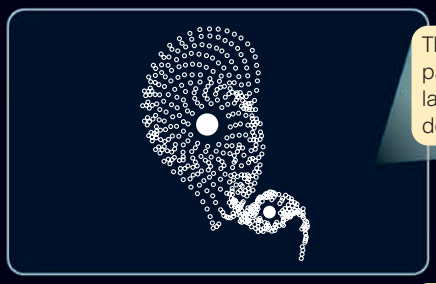
Illustration design by Michael A. Seeds

1a When a galaxy swings past a massive object such as another galaxy, tides are severe. Stars near the massive object try to move in smaller, faster orbits whereas stars farther from the massive object follow larger, slower orbits. Such tides can distort a galaxy or even rip it apart.

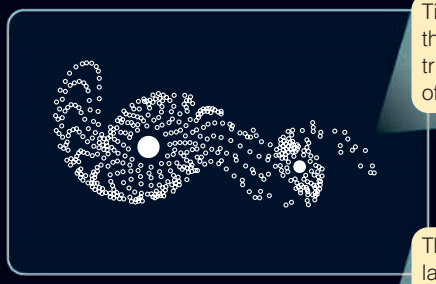
Galaxy interactions can stimulate the formation of spiral arms.



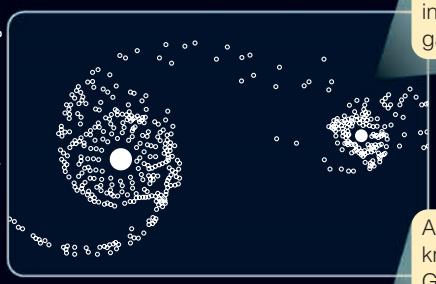
In this computer model, two uniform disk galaxies pass near each other.



The small galaxy passes behind the larger galaxy so they do not actually collide.



Tidal forces deform the galaxies and trigger the formation of spiral arms.

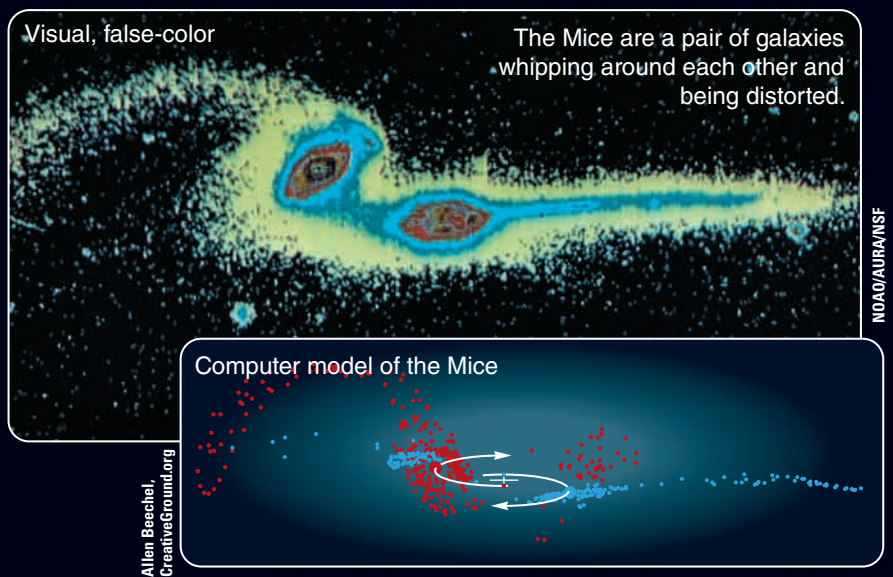


The upper arm of the large galaxy passes in front of the small galaxy.

Allen Beechel, CreativeGround.org

A photo of the well-known Whirlpool Galaxy resembles the computer model.

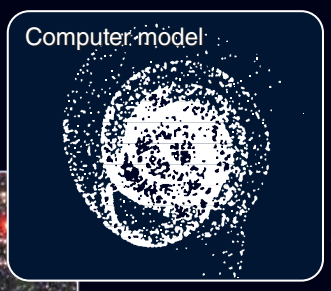
NOAO/AURA/NSF


Visual, false-color
The Mice are a pair of galaxies whipping around each other and being distorted.
Computer model of the Mice

NOAO/AURA/NSF

Allen Beechel, CreativeGround.org

1b The merger of galaxies is called **galactic cannibalism**. Models show that merging galaxies spiral around their common center of mass whereas tides rip stars away and form shells.



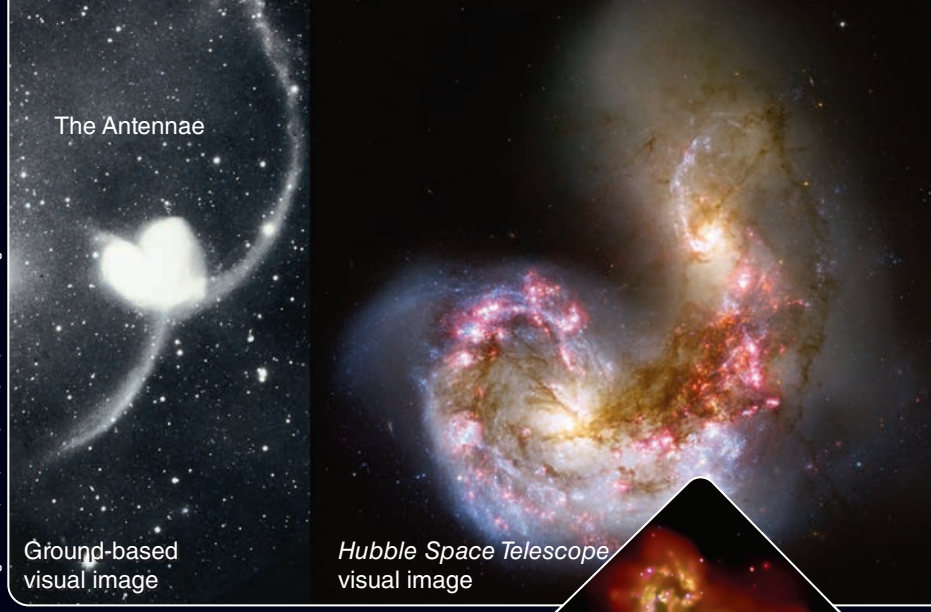
F. Schweizer and A. Toomre

CTIO/NOAO/AURA/NSF/E. Peng and H. Ford (JHU), K. Freeman (ANU), R. White (STScI)

Such shells have been found around elliptical galaxies such as NGC 5128. It is peculiar in many ways and even has a belt of dusty gas. The shells revealed in this enhanced image are evidence that the giant galaxy has cannibalized at least one smaller galaxy. The giant galaxy itself may be the result of the merger of two large galaxies.

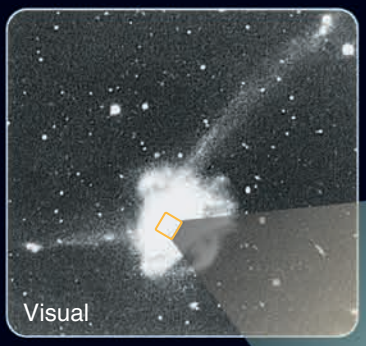
2 The collision of two galaxies can trigger firestorms of star formation as gas clouds are compressed. Galaxies NGC 4038 and 4039 have been known for years as the Antennae because the long tails visible in Earth-based photos resemble the antennae of an insect. *Hubble Space Telescope* images reveal that the two galaxies are blazing with star formation. Roughly a thousand massive star clusters have been born.

Spectra show that the Antennae Galaxies are 10 to 20 times richer in elements such as magnesium and silicon than the Milky Way. Such metals are produced by nucleosynthesis in massive stars and spread back into space by subsequent supernova explosions.



Left: F. Schweizer (CIW) and B. Whitmore (STScI); Right: NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team

3 Evidence of past galaxy mergers shows up in the motions inside some galaxies. NGC 7251 is a highly distorted galaxy with tidal tails in this ground-based image.

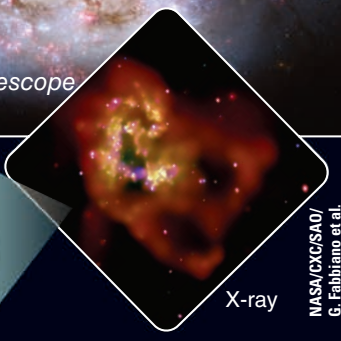


F. Schweizer (CIW) and B. Whitmore (STScI)

This *Hubble Space Telescope* image of the core of the galaxy reveals a small spiral spinning backward in the heart of the larger galaxy.

This counter rotation suggests that NGC 7251 is the remains of two oppositely rotating galaxies that merged about a billion years ago.

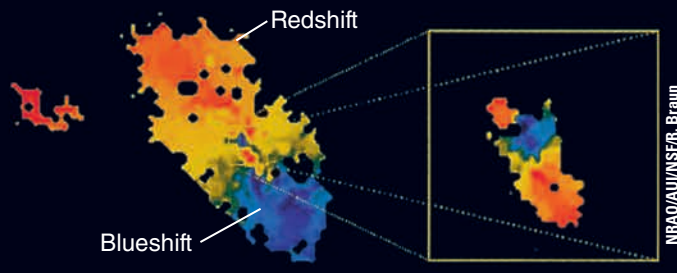
An X-ray image of the Antennae shows clouds of very hot gas heated by supernovae exploding 30 times more often than in our own galaxy.



NASA/CXC/SAO/ G. Fabbiano et al.

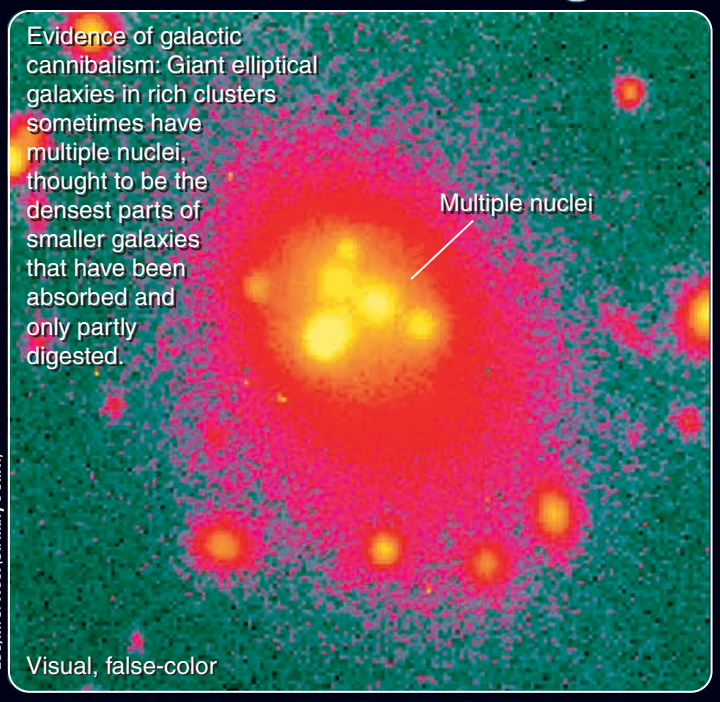
3a Radio evidence of past mergers: Doppler shifts reveal the rotation of the spiral galaxy M64. The upper part of the galaxy has a redshift and is moving away from Earth, and the bottom part of the galaxy has a blueshift and is approaching. A radio map of the core of the galaxy reveals that it is rotating backward. This suggests a merger long ago between two galaxies that rotate in opposite directions.

Rotation of galaxy M64



NRAO/AUI/NSF/R. Braun

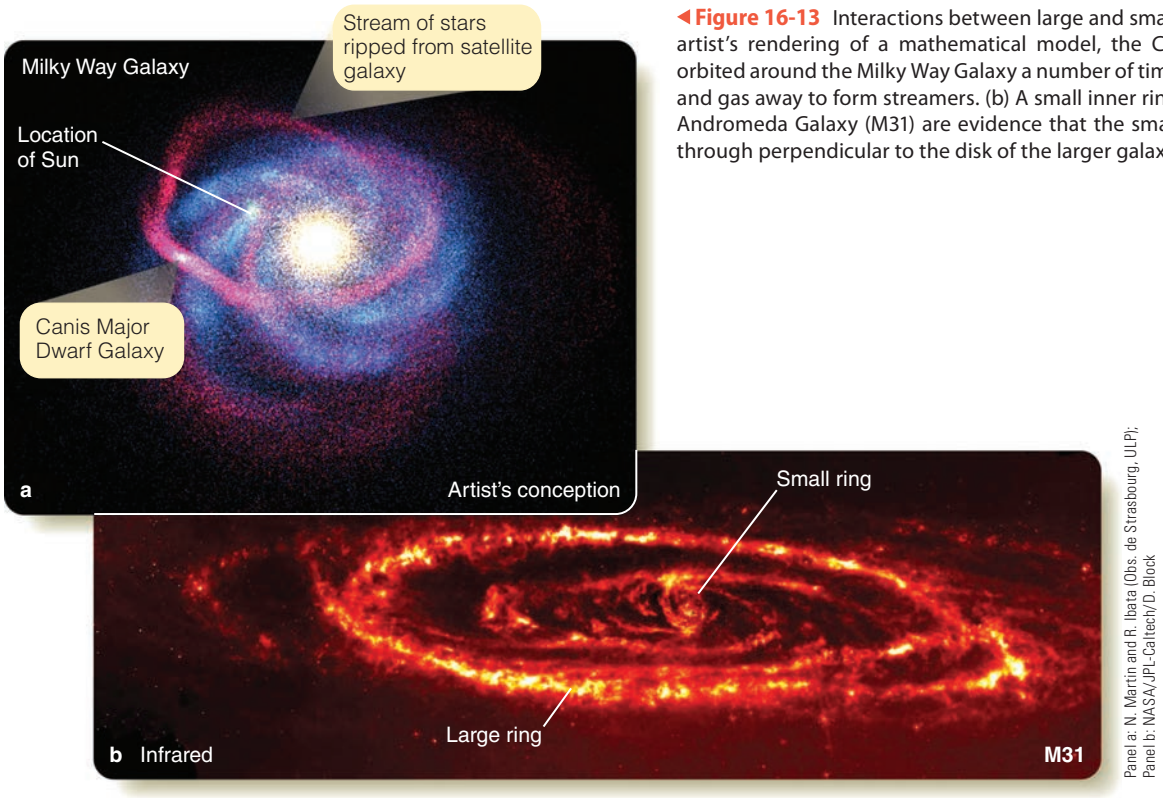
Evidence of galactic cannibalism: Giant elliptical galaxies in rich clusters sometimes have multiple nuclei, thought to be the densest parts of smaller galaxies that have been absorbed and only partly digested.



ESO/M. J. West (St. Mary's Univ.)

4 The Cartwheel Galaxy pictured below was once a normal spiral or lenticular galaxy but is now a **ring galaxy**. One of its smaller companions has plunged through at high speed almost perpendicular at the Cartwheel's disk. That has triggered a wave of star formation, and the more massive stars have exploded leaving behind black holes and neutron stars. Some of those are in X-ray binaries, and that makes the outer ring bright in X-rays.

NASA/JPL-Caltech/P. N. Appleton (SSC)



Panel a: N. Martin and R. Ibata (Obs. de Strasbourg, ULP); Panel b: NASA/JPL-Caltech/D. Block

◀ **Figure 16-13** Interactions between large and small galaxies: (a) As shown in this artist's rendering of a mathematical model, the Canis Major Dwarf Galaxy has orbited around the Milky Way Galaxy a number of times, and tides have ripped stars and gas away to form streamers. (b) A small inner ring and a large outer ring in the Andromeda Galaxy (M31) are evidence that the small satellite galaxy M32 passed through perpendicular to the disk of the larger galaxy about 200 million years ago.

Infrared observations of the nearby Andromeda Galaxy show that it contains two dusty rings, a small inner ring and a larger outer ring. These appear to have been produced when the dwarf elliptical galaxy M32 plunged through the galaxy, moving nearly perpendicular to the disk. You can see the rings in Figure 16-13b.

You have seen the evidence that collisions and mergers between galaxies are common and can produce dramatic changes in the structure of the galaxies; now you are ready to evaluate a hypothesis that may explain how galaxies form and evolve.

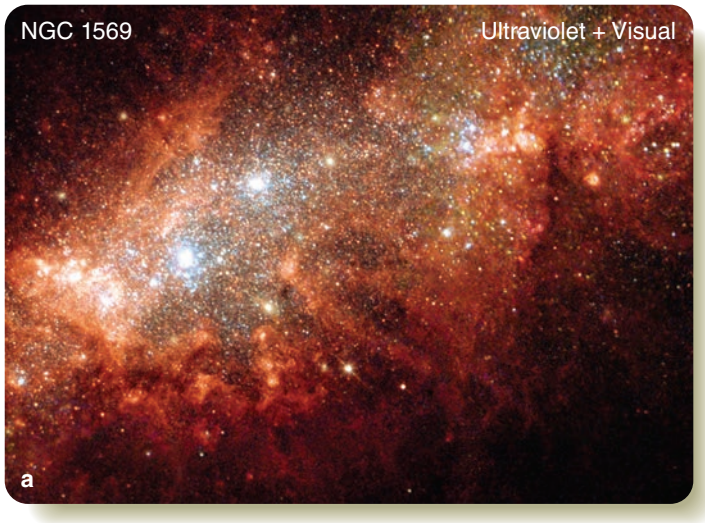
16-3c Origin and Evolution of Galaxies

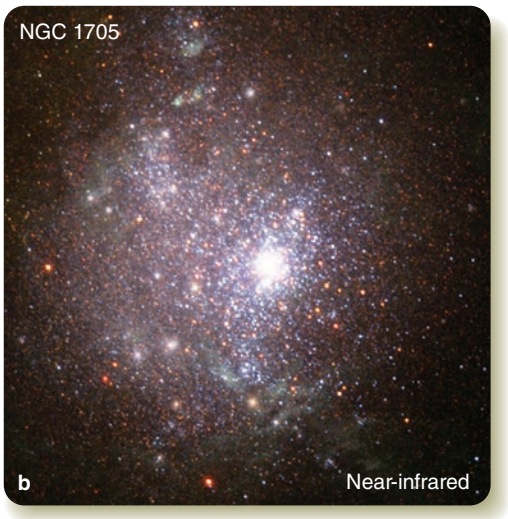
The test of a complete scientific understanding of some type of object is whether you can put all the evidence and theory together to tell their history. Just a few decades ago, it would have been impossible to describe the origin and evolution of galaxies, but the evidence from space telescopes and new-generation telescopes on Earth, combined with advances in computer modeling and theory, allows astronomers to outline the stories of the galaxies.

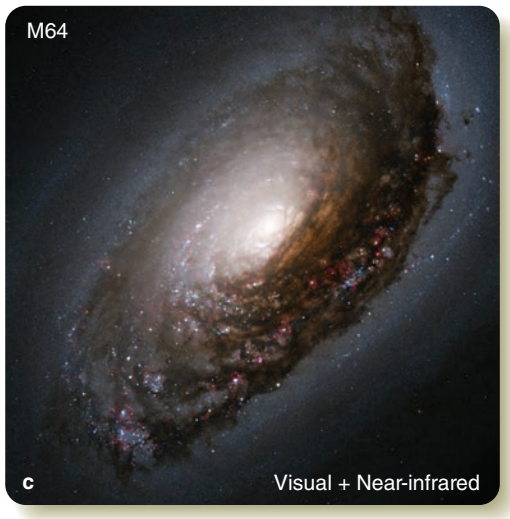
Before you begin, you should eliminate a few older ideas immediately. Hubble and other astronomers who were among the first to study galaxies hypothesized that galaxies normally evolve from one type to another. But now astronomers know that an elliptical galaxy cannot become a spiral galaxy or an irregular galaxy because ellipticals contain almost no gas and dust from which to make new stars. Also, spiral and irregular galaxies contain old stars, so those types of galaxies are not "young." The galaxy classes tell you something important, but evidently an isolated galaxy does not change from one class to another any more than a cat can change into a dog.

Another old hypothesis was that a galaxy that formed from a rapidly rotating protogalaxy gas cloud would have lots of angular momentum and would contract slowly, forming a disk-shaped spiral galaxy. A protogalaxy cloud that rotated less rapidly would contract faster, form stars quickly, use up all of its gas and dust, and become an elliptical galaxy. This hypothesis was consistent with the original hypothesis that the Milky Way and other galaxies formed from the top down—from individual large clouds of gas. That idea is described in Section 15-5e as the top-down hypothesis. Modern evidence clearly shows that large galaxies such as the Milky Way form partly from the bottom up—from the gradual accumulation of smaller clouds of gas and stars and, in some cases, entire small galaxies. Collisions and mergers apparently dominate the histories of large galaxies.

Elliptical galaxies appear especially to be the product of galaxy collisions and mergers. They are devoid of gas and dust because it was used up in the rapid star formation triggered by the interaction. In fact, astronomers see many **starburst galaxies** that are very luminous in the infrared because a recent collision has triggered a burst of star formation that is heating their interstellar dust clouds (**Figure 16-14**), which reradiate the energy in the infrared. Some of those galaxies are a hundred times more luminous than







Panel a: NASA/ESA/STScI/AURA/NSF/P. Anders et al.; Panels b and c: NASA/ESA/STScI/AURA/NSF/Hubble Heritage Team

◀ **Figure 16-14** Rapid star formation: (a) NGC 1569 is a starburst galaxy filled with clouds of young stars containing many supernova remnants. At least some starbursts are triggered by interactions between galaxies. (b) The dwarf irregular galaxy NGC 1705 began a burst of star formation about 25 million years ago. (c) The inner parts of M64, known as the "Black Eye Galaxy," are filled with dust produced by rapid star formation. Radio observations show that the inner part of the galaxy rotates backward compared to the outer part of the galaxy, a product of a merger. Where the counter-rotating parts of the galaxy collide, star formation is stimulated.

our Milky Way Galaxy but so deeply shrouded in dust that they are very dim at visible wavelengths. Such **ultraluminous infrared galaxies (ULIRGs)** often show evidence of tidal tails and are probably the result of the merger of three or more galaxies that triggered firestorms of star formation and generated tremendous clouds of dust that condensed in supergiant stellar winds and supernova explosion remnants. Supernova explosions can also serve to expel interstellar material from galaxies (Figure 15-21).

The Antennae Galaxies (Figure 2 in Concept Art: Interacting Galaxies) contain more than 15 billion $M_{\odot}$ of hydrogen gas and will become a starburst galaxy as the ongoing merger continues to trigger rapid star formation. Model calculations indicate that when those galaxies use up the last of their interstellar gas making stars, they will probably merge to form a single elliptical galaxy.

In this way, a few collisions and mergers could leave a galaxy with no gas and dust from which to make new stars and also could scramble the orbits of the remaining stars to produce an elliptical shape. Astronomers now suspect that most large elliptical galaxies are formed by the merger of at least two or three smaller galaxies.

In contrast, spirals evidently have never suffered major collisions. Their thin disks are delicate and would be destroyed by the tidal forces generated during a collision with a massive galaxy. Also, they retain plenty of gas and dust and continue making stars, so they have never experienced a major starburst triggered by a merger.

Of course, spiral galaxies can safely cannibalize smaller galaxies with no ill effects. The small galaxies do not generate strong tides to distort a full-sized galaxy. In Section 15-5e, the bottom-up model for galaxy formation implied that our Galaxy has absorbed many smaller galaxies, and some astronomers suspect that much of the halo consists of the remains of cannibalized galaxies. Evidently our Milky Way Galaxy has not collided with a galaxy of size comparable to itself, so far. Such a collision would trigger rapid star formation, use up gas and dust, scramble the orbits of stars, and destroy the thin disk. However, observations

indicate that our Galaxy will merge with the approaching Andromeda Galaxy in a few billion years, and the final result probably will be a large elliptical galaxy.

Just as people do, spiral galaxies may need some interaction to fully develop. Only one in 10,000 galaxies is isolated from its neighbors; most of these are disk galaxies, and many are flocculent, without a strong two-armed spiral pattern. Some isolated galaxies are rich in gas and dust but have little star formation. All of this suggests, as you learned in Section 15-3c, that continuous gentle gravitational interactions with neighboring galaxies may be necessary to stimulate spiral density waves, strong spiral structure, and ongoing star formation.

Earlier in this chapter, the Virgo cluster was described as a rich cluster, so you may find it surprising to learn that it contains lots of spiral galaxies. Rich clusters are supposed to contain few spirals. Astronomers hypothesize that this cluster is so massive that the galaxies orbit the center of mass of the cluster at high enough speeds that they rush past each other too quickly to interact strongly. The galaxies are whittled down a bit, but many are still disk galaxies with spiral patterns.

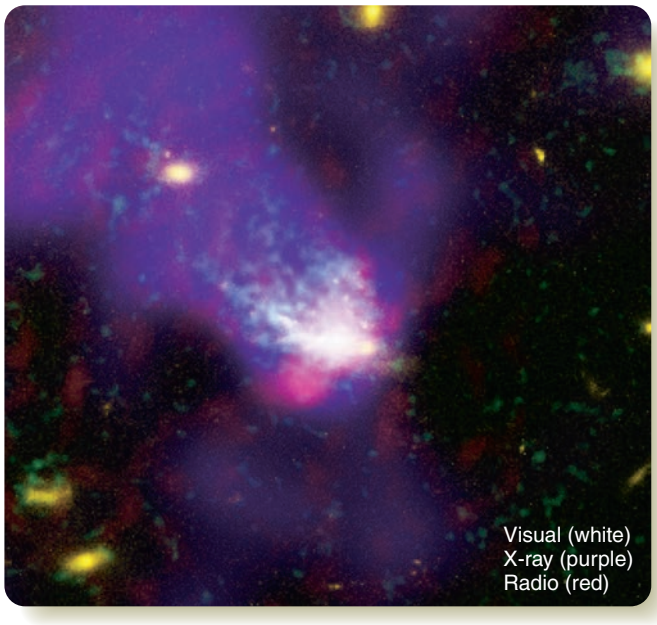
Barred spiral galaxies may also be the products of tidal interactions. Mathematical models show that bars are not stable and eventually dissipate. Tidal interactions with other galaxies may regenerate the bars. Approximately two-thirds of all spiral galaxies, including our Milky Way Galaxy, have bars, suggesting that these tidal interactions are common.

Other processes can alter galaxies. The S0 galaxies retain a disk shape but lack gas and dust. Although they could have lost gas and dust during starburst episodes, they may also have lost it as they orbited through the gas in their respective clusters. A galaxy moving rapidly through the thin gas filling a cluster would encounter a tremendous wind that could blow away the galaxy's gas and dust. In this way, a disk-shaped spiral galaxy could be reduced to an S0 galaxy. For example, X-ray observations show that the Coma cluster contains thin, hot gas between the galaxies (Figure 16-9), and astronomers have located, in a similar cluster, a galaxy in the act of plunging through such gas and being stripped of its gas and dust (**Figure 16-15**).

Small galaxies may be produced in a number of ways. The dwarf ellipticals are too small to be produced by mergers, but they could be fragments of galaxies ripped free during interactions. Another possibility is that dwarf ellipticals are small galaxies that have plunged through the gas in a cluster of galaxies and lost their gas and dust. In contrast, the irregular galaxies may be small fragments splashed from larger galaxies during collisions that retain enough gas and dust to continue forming stars.

Other factors must influence the evolution of galaxies. Cool gas clouds may fall into galaxies from their surroundings and add material for star formation. These processes are just beginning to be understood.

A comprehensive theory helps you understand how nature works, and astronomers are beginning to understand the exciting and complex story of the galaxies. Nevertheless, it is already clear that galaxy evolution bears some resemblance to a pie-throwing contest.



NASA, W. Keel (Univ. of Alabama), F. Owen (NRAO), M. Ledlow (Gemini Obs.), and D. Wang (Univ. of Massachusetts)

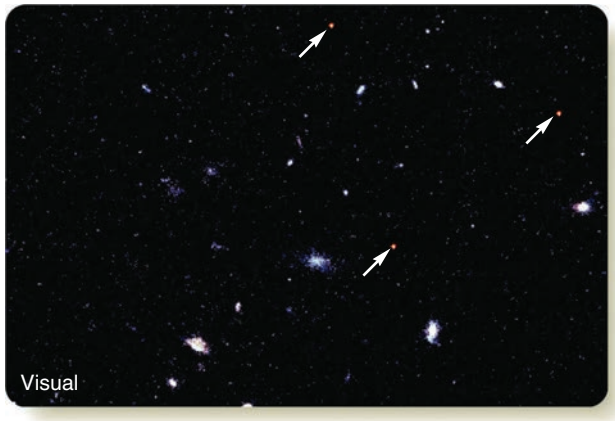
▲ **Figure 16-15** The distorted galaxy C153 is orbiting through the thin gas in its home cluster of galaxies at 2000 km/s (4.5 million mph). At that speed, it is subjected to a tremendous wind stripping gas out of the galaxy in a trail 200,000 ly long. Such galaxies could quickly lose almost all of their gas and dust.

16-3d The Farthest Galaxies

Observations made with the largest and most sophisticated telescopes are taking astronomers back to the age of galaxy formation. At great distances, the look-back time is so great that the Universe can be seen as it was soon after galaxies began to form. There were more spirals then and fewer ellipticals. On the whole, galaxies long ago were more compact and more irregular than they are now. The observations also show that galaxies were closer together long ago; about a third of all distant galaxies are in close pairs, but only 7 percent of nearby galaxies are in pairs. The observational evidence clearly supports the hypothesis that galaxies have evolved by collision and merger.

At great distances, astronomers see tremendous numbers of small, blue, irregularly shaped objects that have been called **blue dwarf galaxies**. The blue color indicates that they are rapidly forming stars, but the role of these blue dwarfs in the formation of galaxies is not clear. Blue dwarf galaxies are not found at small look-back times, so they no longer exist in the present Universe. They may be clouds of gas and stars that were absorbed long ago in the formation of larger galaxies, or they may have used up their gas and dust and faded into obscurity.

At the limits of the largest telescopes, astronomers see faint red galaxies (**Figure 16-16**). They look red from Earth because of their great redshifts. The floods of ultraviolet light emitted by these star-forming galaxies have been shifted far into the red part



NASA/ESA/STScI/AURA/NSF/ H.-J. Yan, R. Windhorst, and S. Cohen (Arizona State Univ.)

▲ **Figure 16-16** Arrows point to three very distant red galaxies. The look-back time to these galaxies is so large that they appear as they were when the Universe was only a billion years old. Such highly redshifted galaxies are understood to have been among the first to form stars and begin shining after the beginning of the Universe about 14 billion years ago.

of the spectrum. Their look-back times are so great that they appear as they were when the Universe was only a billion years old. These galaxies seem to be among the first to begin shining after the beginning of the Universe, a story told in Chapter 18.

Practicing **SCIENCE**

How did elliptical galaxies get their shape? Being able to figure out the histories of objects in the natural world is a goal that lies at the heart of science.

A growing body of evidence suggests that elliptical galaxies have been subject to collisions in their past and that spiral galaxies have not. During collisions, a galaxy can be driven to use up its gas in a burst of star formation, and the resulting supernova explosions can help drive interstellar material out of the galaxy. This explains why elliptical galaxies now contain little star-making material. The beautiful disk typical of spiral galaxies is very orderly, with all the stars following similar orbits. When galaxies collide, the stellar orbits get scrambled, and an orderly disk galaxy could be converted into a chaotic swarm of stars typical of elliptical galaxies. It seems likely that elliptical galaxies have had more complex histories than spiral galaxies have had.

When scientists compose the story of a natural object, they use evidence to make the story as true as it can be. Try to work out another story, based on observations of our own Milky Way Galaxy. ***How do spiral galaxies get their shape?***

WHAT ARE WE?

Small and Proud

Do you feel insignificant yet? You are riding a small planet orbiting a humdrum star that is just one of at least 200 billion stars in the Milky Way Galaxy, and you have just learned that there are more than 100 billion galaxies visible with existing telescopes.

When you look at galaxies, you are looking across voids deeper than human imagination. You can express such distances with numbers and say a certain galaxy is 5 billion ly from Earth, but the distance is truly beyond human comprehension. Furthermore, looking at distant galaxies also leads you back in time. You see distant galaxies as they were billions of years ago. The realm of the galaxies is deep space and deep time.

Some people say astronomy makes them feel humble, but before you agree, consider that you can feel small without feeling humble. Size is not necessarily the same thing as significance. We humans live out our little lives on our little planet, but our tiny brains are able to figure out some of the most profound mysteries of the Universe. We are exploring deep space and deep time and coming to understand what galaxies are and how they evolve. Most of all, we humans are beginning to understand what we are. That's something to be proud of.

Study and Review

SUMMARY

16-1 The Family of Galaxies

- Through 19th-century telescopes, most galaxies looked like hazy **spiral nebulae**. Some astronomers hypothesized that they are other star systems called island universes, but others hypothesized that these spiral nebulae are clouds of gas within the Milky Way star system. The controversy culminated in the **Shapley–Curtis debate** in 1920. A few years later, with the construction of larger telescopes, astronomers identified distant galaxies rivaling the Milky Way Galaxy in size.
- Astronomers divide galaxies into three major types—**elliptical, spiral**, and **irregular**—with subtypes specifying details of the galaxy's shape. **Lenticular galaxies** (class S0) are disk galaxies with no spiral arms and, usually, little interstellar gas or dust. About two-thirds of spirals, including our Milky Way Galaxy, are **barred**

spiral galaxies. Irregular galaxies have no obvious shape but contain gas and dust and support star formation.

16-2 Measuring the Properties of Galaxies

- Galaxies are so distant astronomers measure their distances in **megaparsecs (Mpc)**—millions of parsecs. Astronomers find the distance to galaxies using **distance indicators**, such as Cepheid variable stars, that are objects of known luminosity. By calibrating several types of distance indicators using galaxies of known distance, astronomers have built a **distance scale**. When astronomers look at a distant galaxy, they see the galaxy as it was when it emitted the light that is now reaching Earth as a result of **look-back time**.
- According to the **Hubble law**, the apparent velocity of recession of a galaxy equals its distance times the **Hubble constant**. Astronomers can estimate the distance to a galaxy by observing the galaxy's redshift, calculating the galaxy's apparent velocity of recession, and then dividing by the Hubble constant.
- Once the distance to a galaxy is known, the galaxy's diameter can be found from the small-angle formula; its luminosity can be found from the magnitude–distance relation.
- Astronomers measure the masses of galaxies in two basic ways. The orbital motion of a disk galaxy's stars can be measured using a spectrograph, and the **rotation curve method** can then be used to calculate the galaxy's mass. The **velocity dispersion method** depends on measuring the random velocities of the stars in an elliptical galaxy to find the total mass of the galaxy. The **cluster method** uses the random velocities of the galaxies in a cluster to find the total mass of the cluster, analogous to the velocity dispersion method for individual elliptical galaxies.
- Stars and gas clouds near the centers of large galaxies are following small orbits at high velocities, suggesting the presence of supermassive black holes in the centers of most such galaxies. The mass of a galaxy's central supermassive black hole is proportional to the mass of its central bulge. This indicates that those black holes must have formed when the galaxy bulges formed.
- Observations of individual galaxies show that galaxies contain at least 10 times more dark matter than visible matter.
- The hot gas held inside some clusters of galaxies and the **gravitational lensing** caused by the mass of galaxy clusters reveal that the clusters must be much more massive than can be accounted for by just the visible matter—further evidence of dark matter.

16-3 The Evolution of Galaxies

- **Rich clusters** of galaxies contain thousands of galaxies, with few spirals and more ellipticals. **Poor clusters** of galaxies contain few galaxies with a larger proportion of spirals. This is evidence that galaxies evolve by collisions and mergers. Our galaxy is located in a poor cluster known as the **Local Group**.
- When galaxies collide, tides twist and distort the galaxies' shapes and can produce **tidal tails**. Large galaxies can absorb smaller galaxies in what is called **galactic cannibalism**. **Ring galaxies** are produced by high-speed collisions in which a small galaxy plunges through a larger galaxy perpendicularly to the larger galaxy's disk.
- The compression of gas clouds can trigger rapid star formation, producing **starburst galaxies**. The rapid star formation can produce lots of dust, which is warmed by the stars to emit infrared radiation, making the galaxy an **ultraluminous infrared galaxy (ULIRG)**. At great distances and great look-back times, the largest telescopes reveal that there were more spirals and fewer elliptical galaxies when the Universe was young; also, galaxies were smaller, more irregular, and closer together.
- At the largest distances, astronomers find small, irregular clouds of blue stars called **blue dwarf galaxies** that may be the original components that combined to begin forming large galaxies.

REVIEW QUESTIONS

1. If a civilization lived on an exoplanet in an E0 galaxy, do you think it would have a "Milky Way" band of starlight in its sky? Why or why not?
2. Of the nearby galaxies, which is the most common type? The least?
3. What is the type and subtype of a galaxy that is shaped like a brick, has lots of gas and dust, and contains hot blue stars in its wide spiral arms?
4. What is the classification of a galaxy that is round, with no spiral arms and no disk component?
5. What is the classification of the Milky Way Galaxy?
6. Which are more common, barred or nonbarred spiral galaxies?
7. Why can't isolated galaxies evolve from elliptical to spiral? From spiral to elliptical?
8. Stars collide often in galaxies, whereas galaxies collide rarely. True or false? How do you know?
9. If all elliptical galaxies have three different diameters, would you ever see an elliptical galaxy with a circular outline in a photograph? Explain your answer. (*Hint:* Can a squashed bread loaf, with three different dimensions, ever cast a circular shadow no matter how it is oriented?)
10. What is the difference between an Sa and an Sb galaxy? Between an S0 and an Sa galaxy? Between an Sb and an SBb galaxy? Between an E7 and an E0 galaxy? Between an E0 and an S0 galaxy?
11. Would a 100-watt lightbulb make a good distance indicator? Why or why not?
12. Why wouldn't white dwarf stars make good distance indicators?
13. Would a type I supernova be a good distance indicator? Would a type II supernova be a good distance indicator? How do you know?
14. Why isn't the look-back time important among nearby galaxies?
15. The distance to M31, the Andromeda Galaxy, is about 2,500,000 ly. If you can see M31 in the night sky tonight, when did the light you are receiving leave the galaxy? How far away is M31 in Mpc?
16. How would you find the distance to a nearby star in order to help calibrate the distance ladder?
17. How would you find the distance to a nearby galaxy in order to help calibrate the distance ladder?
18. How would you find the distance to a very far galaxy?
19. Name three basic properties of galaxies and explain why these are considered independent (that is, the properties do not depend on each other).
20. Explain how the rotation curve method of finding a galaxy's mass is similar to the method used to find the masses of binary stars.
21. Explain how the Hubble law allows you to estimate distances to galaxies.
22. What is the percentage range of galaxy diameters relative to the diameter of the Milky Way Galaxy?
23. What is the percentage range of galaxy masses relative to the mass of the Milky Way Galaxy?
24. Examine Table 16-1. What connection is there, if any, between galaxy mass, diameter, and luminosity, and galaxy type? Explain your answer.
25. Of 100 percent of all matter in galaxies, approximately what fraction is visible matter?
26. Did the Milky Way Galaxy's central supermassive black hole form before, at the same time as, or after the galaxy formed? How do you know?

27. Is the Milky Way Galaxy part of a rich or poor cluster? How do you know? What is the name of our cluster?
28. How might collisions affect the shape of galaxies?
29. What evidence can you provide that galactic cannibalism really happens?
30. Describe the future evolution of a galaxy that astronomers now see as a starburst galaxy. What will happen to the interstellar medium in this galaxy?
31. If two equal-size spiral galaxies were to collide, what would happen?
32. Why does the gas within a cluster of galaxies sometimes help determine the nature of the galaxies in that cluster?
33. Why are blue dwarf galaxies blue? Where are blue dwarf galaxies located relative to us?
34. **How Do We Know?** Classification helped Darwin understand how creatures evolve. How has classification helped astronomers understand galaxies and how they evolve?

ACTIVE INQUIRY QUESTIONS

1. From what you know about star formation and the evolution of galaxies, do you think irregular galaxies should be bright or faint in the infrared? Why or why not? What about starburst galaxies? What about elliptical galaxies?
2. From what you know about evolution of galaxies, cannibalism, and the Local Group, make a hypothesis about the shape of the Milky Way Galaxy in 3 billion years.
3. Imagine that you could observe a few gas clouds at a large look-back time and you saw that these clouds are just beginning to form one of the first galaxies. Suppose also that from your observations you found that the gas in those clouds is metal rich. How would that affect your understanding of galaxy formation?
4. A grand-design spiral galaxy is 250 million ly from Earth. If you saw the galaxy in the night sky tonight, when did the light leave the galaxy? Were dinosaurs present on Earth at the time when the light left? If you take an image of the galaxy tonight, will it depict what the galaxy would look like right now, in a general sense? In detail? (*Hint:* See Section 1-2.)

PROBLEMS

1. Measure the dimensions a and b of Galaxy M87 in Concept Art: Galaxy Classifications. Determine the shape index of M87 using your measurements. Is this a highly elliptical or nearly spherical galaxy? List your measurements and show your work.
2. If the shape index is zero for a particular elliptical galaxy, what is the major axis length a relative to the minor axis length b? Is this a highly elliptical or nearly spherical galaxy?
3. If a galaxy contains a type I (classical) Cepheid that has a pulsation period of 30 days and an apparent magnitude of +20, what is the distance to the galaxy? (*Hints:* Refer to Figure 12-14, and use the magnitude–distance formula, Eq. 9-2b.)
4. RR Lyrae has a 0.65-day period, and $m_V = 7.2$. What is its distance? (*Hints:* Refer to Figure 12-14 and use the magnitude–distance formula, Eq. 9-2b.)
5. Assume an average globular cluster is 25 pc in diameter. You observe a galaxy that contains globular clusters that are 2 arc seconds in angular diameter. How far away is the galaxy? (*Hint:* Use the small-angle formula, Eq. 3-1.)
6. If you find a galaxy that has an angular diameter of 1200 arc seconds and you measure its distance to be 1 Mpc, what is its true diameter? (*Hint:* Use the small-angle formula, Eq. 3-1.)
7. You find a Cepheid variable in another galaxy and you determine the galaxy's distance to be 12 Mpc. Calculate the galaxy's apparent radial velocity V using the Hubble law.
8. If a galaxy has an apparent radial velocity $V = 2000$ km/s, use the Hubble law to calculate the distance to the galaxy.
9. If a galaxy contains a supernova that has an apparent magnitude of +17 at its brightest, how far away is the galaxy? Assume that the absolute magnitude of the supernova is −19. (*Hint:* Use the magnitude–distance formula, Eq. 9-2b.)
10. Hubble's first determination of what is now called the Hubble constant was too high because the calibration of Cepheid variable stars he used was not very good. Use his original diagram shown in Figure 16-5 to estimate his first determination of the constant.
11. Study Figure 16-7. Compare this rotation curve to that of Keplerian motion (that is, Kepler's third law) of the planets within the Solar System. What does the difference in the shape of the graph tell you? How do you know?
12. Suppose you found a spiral galaxy in which the outermost stars have orbital velocities of 150 km/s. If the radius of the galaxy is 4.0 kpc, what is the orbital period of those stars? (*Note:* 1 pc $= 3.1 \times 10^{13}$ km and 1 yr $= 3.2 \times 10^{7}$ s.)
13. If you find a galaxy that is the same size and mass as our Milky Way Galaxy, what orbital velocity would a small satellite galaxy have that is orbiting 50 kpc from the center of the larger galaxy? On what assumption or assumptions does this result depend? (*Hint:* See the calculation of the Milky Way Galaxy's mass in Chapter 15.)
14. Find the orbital period of the satellite galaxy described in Problem 13.
15. A galaxy has been found that is 5 kpc in radius and whose outer stars orbit the center with a period of 200 million years. What is the estimated mass of the galaxy? On what assumption or assumptions does this result depend? (*Hint:* See the calculation of the Milky Way Galaxy's mass in Chapter 15.)
16. Among the globular clusters orbiting a distant galaxy, the fastest is traveling at 420 km/s and is located 11 kpc from the center of the galaxy. Assuming the globular cluster is not escaping but is just barely gravitationally bound to the galaxy, what is the mass of the galaxy? (*Hint:* The galaxy had a slightly faster globular cluster, but it escaped some time ago.)

SENSE OF PROPORTION

1. Arrange the following in order of increasing size: our Milky Way Galaxy; a globular star cluster; a pair of radio lobes around an active galaxy; a giant elliptical galaxy; the Solar System; the Local Group galaxy cluster
2. If the diameter of the Milky Way Galaxy's visible disk, 80,000 ly, is represented in a model by a dinner plate with a diameter of 10 inches, what is the model distance to galaxy M31, 2.6 million light-years away? What is the model distance to the Virgo galaxy cluster, 16 megaparsecs away? (Convert answers to feet.)

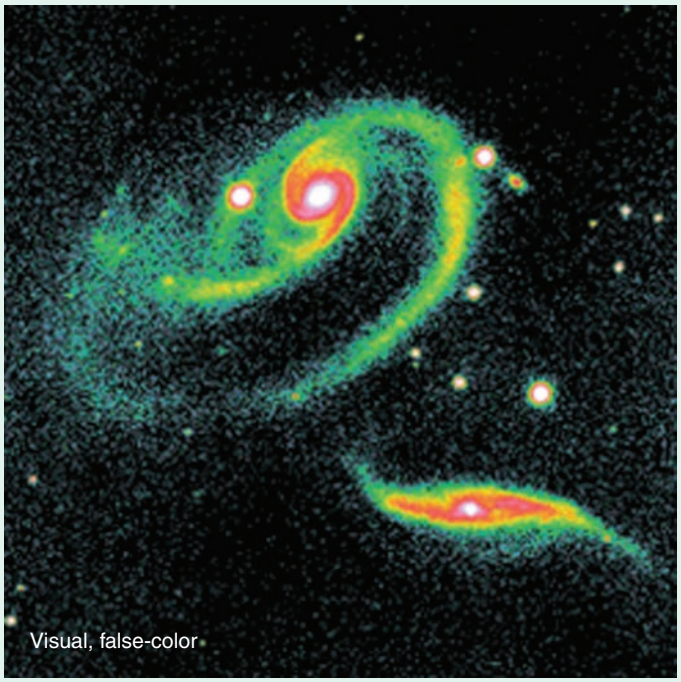
LEARNING TO LOOK

1. Study Figure 16-2. Which galaxies do you suppose are the nearest ones, and which are the farthest? How is your estimate based on your own calibration of galaxy size and luminosity?
2. Study the visual wavelength image of the galaxy shown in Learning to Look question 15-4. What is that galaxy's class? What is its subclass? How do you know?
3. Study the visual wavelength image of the galaxy shown in Learning to Look question 15-5. What is that galaxy's class? What is its subclass? How do you know?

4. Study Galaxies M87 and NGC 2997 in Concept Art: Galaxy Classifications. Which one has OB associations? Which one has more gas and dust? Which one has more stars? How do you know?
5. This image of Galaxy M33 has emission by ionized hydrogen enhanced as bright pink. Discuss the location of these clouds of gas and explain how this location provides important evidence toward understanding spiral arms.

NOAO/NSF/WIYN/P. Massey (Lowell Obs.), N. King (STScI), S. Holmes (College of Charleston), G. Jacoby (NOAO)

6. In the image shown here, you see two interacting galaxies; one is nearly face-on, and the other is nearly edge-on. Discuss the shapes of these galaxies and describe what is happening.



WIYN/NOAO/AURA/NSF

Active Galaxies and Supermassive Black Holes

17

X-ray: NASA/CXC/SAO/KIPAC/N. Werner, E. Million et al. Radio: NRAO/AUI/NSF/F. N. Owen

▲ Jets and outflows from the nucleus of giant elliptical galaxy M87 are revealed by their X-ray emission. Those flows are pushing into the cooler gas detected by a radio telescope, with the resulting spherical shock wave observed expanding outward. The galaxy as seen at visual wavelengths would entirely fill this frame.

In the previous two chapters, you have explored our own Galaxy and other galaxies, and now you are ready to stretch your scientific imagination and study some of the most powerful objects in nature. Supermassive black holes at the centers of galaxies are common but extreme. To study them, you will be combining many of the ideas you have discovered so far to answer four important questions:

- What makes some galaxy nuclei active?
- How do supermassive black holes erupt?
- How did supermassive black holes form and evolve?
- How do supermassive black holes affect their host galaxies and galaxy clusters?

The formation and evolution of supermassive black holes leads your astronomical curiosity outward into space and backward in time to the era of galaxy formation. In the next chapter, you will take the next step and try to understand the birth and evolution of the entire Universe.

> **Somewhere something incredible is waiting to be known.**
>
> CARL SAGAN

ASTRONOMERS HAVE GATHERED EVIDENCE that supermassive black holes containing millions or even billions of solar masses lurk at the centers of most large galaxies, although many of those black holes are "quiet," not actively erupting. Those that do erupt can release tremendous amounts of energy that dwarf the largest supernova explosions. The study of supermassive black holes will introduce you to some of the most extreme conditions in the Universe.

17-1 Active Galactic Nuclei

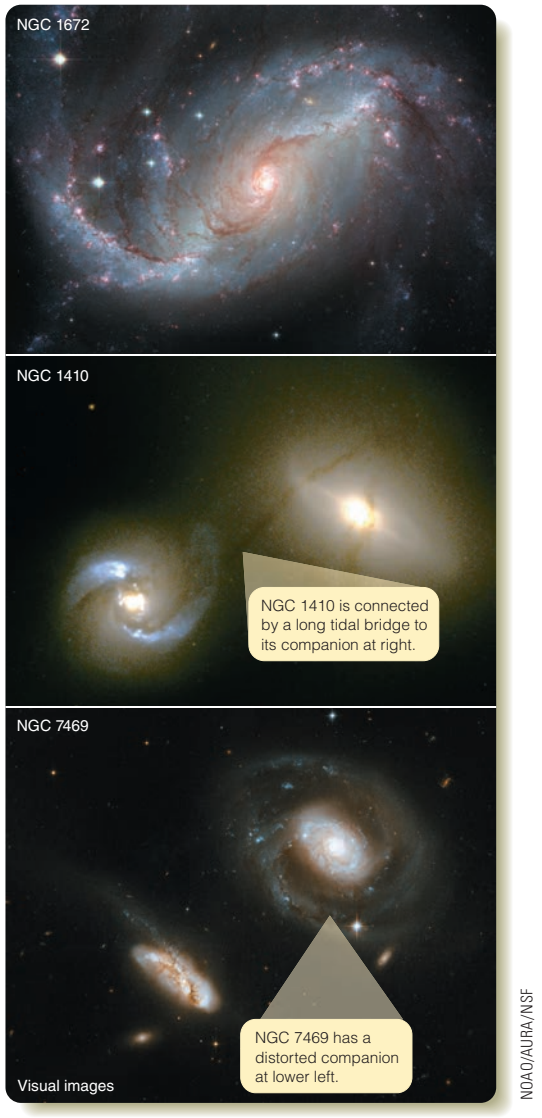
With construction of the first large radio telescopes in the 1950s, astronomers found that some galaxies—dubbed **radio galaxies**—were bright at radio wavelengths. Later, when telescopes went into orbit, these galaxies were found to be emitting energy strongly at other wavelengths as well, and they became known as **active galaxies**. Modern observations show that the energy comes from the nuclei of the galaxies, which are now known as **active galactic nuclei (AGN)**. Only a few percent of galaxies are active, so your first step to understanding this phenomenon is to study those exceptional galaxies.

17-1a Seyfert Galaxies

The first clue that the nuclei of galaxies could be active was found in 1943 by Mount Wilson astronomer Carl Seyfert, who published a comprehensive survey of spiral galaxies. Observing at visual wavelengths, Seyfert found that some spiral galaxies have small, highly luminous nuclei with peculiar spectra. These galaxies are now known as **Seyfert galaxies** (Figure 17-1).

The spectrum of a galaxy is the blended spectra of billions of stars, and consequently weak spectral lines get washed out. Galaxy spectra usually contain only a few of the strongest absorption lines found in stellar spectra. Spectra of Seyfert galaxy nuclei, in comparison, contain broad emission lines. You learned in Section 7-3 that emission lines are produced by a hot, low-density gas, so the gas in the nuclei of Seyfert galaxies must be greatly heated.

The widths of Seyfert galaxy spectral lines are understood to represent Doppler shifts produced by high velocities in the nuclei. Gas approaching Earth produces blueshifted spectral lines, and gas going away produces redshifted lines. The broad lines observed in the spectra of Seyfert galaxies contain light from a combination of gas clouds that are both approaching and receding, with a range of velocities as large as 10,000 km/s. This is about 30 times greater than gas velocities in the centers of normal galaxies. Something violent is happening in the centers of Seyfert galaxies.


NGC 1672
NGC 1410
NGC 1410 is connected by a long tidal bridge to its companion at right.
NGC 7469
NGC 7469 has a distorted companion at lower left.
Visual images
NOAO/AURA/NSF

▲ **Figure 17-1** Seyfert galaxies are spiral galaxies with small, highly luminous nuclei. Some are interacting with nearby companions and appear distorted with tidal tails and bridges.

About 2 percent of spiral galaxies appear to be Seyfert galaxies, and they are classified into two categories. Type 1 Seyfert galaxies are very luminous at X-ray and ultraviolet wavelengths and have the typical broad emission lines with sharp, narrow cores. Rapidly moving gas produces the broad part of the lines, but some lower-velocity gas must also be present to produce the narrow cores. Type 2 Seyfert galaxies have much weaker X-ray emission and have emission lines that are narrower than those of type 1 Seyfert galaxies but still broader than spectral lines observed from normal galaxies.

The brightnesses of Seyfert galaxy nuclei vary rapidly, especially at X-ray wavelengths. A Seyfert nucleus can change its X-ray brightness by a large fraction in only minutes. As you

learned when you studied neutron stars, no object can change its brightness significantly in a time shorter than the time it takes light to cross its diameter. If the Seyfert nucleus can change in a few minutes, then it cannot be more than a few light-minutes in diameter—about the size of Earth's orbit around the Sun. Despite their small size, the nuclei of Seyfert galaxies produce tremendous amounts of energy. The brightest have a hundred times more luminosity than the entire Milky Way Galaxy.

Lots of energy is produced in a very small volume with extremely high temperatures and velocities. What does that remind you of? (*Hint:* Look back to Section 14-2d.) Astronomers conclude that the centers of these galaxies contain supermassive black holes into which matter is flowing from surrounding hot accretion disks.

In the previous chapter, you learned that the stars in the nuclei of most large galaxies are moving with velocities much too high to be explained by only the gravitational influence of those stars. This is understood to be evidence that most large galaxies contain supermassive black holes at their centers. (Note that this is different from the evidence for dark matter. Dark matter also shows its presence gravitationally, but in the outermost regions of galaxies where there is almost no visible material.)

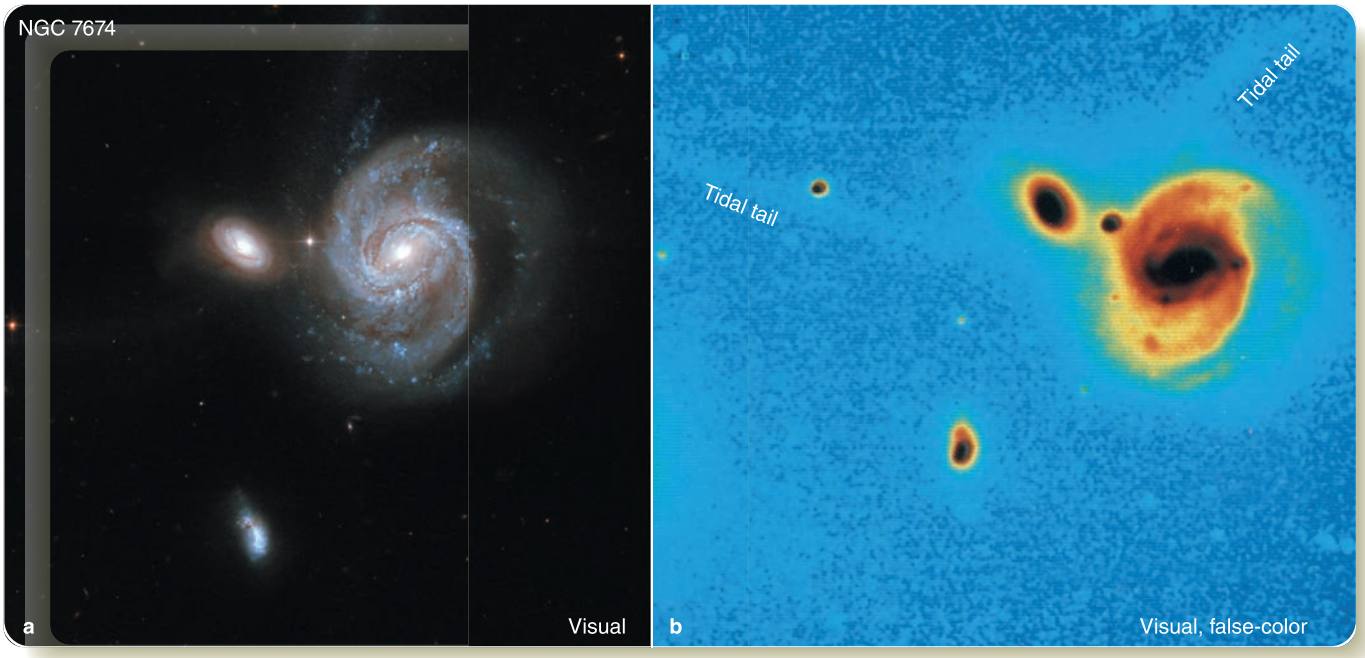
Nevertheless, most galaxies that evidently have supermassive black holes in their nuclei do not show signs of strong activity such as those seen in the few galaxies showing Seyfert activity. The shapes of Seyfert galaxies provide an important clue as to why. About 25 percent of Seyfert galaxies have peculiar shapes, suggesting that they have had tidal interactions with other galaxies (**Figure 17-2**). There is also statistical evidence (**How Do We Know? 17-1**) that Seyfert galaxies are more common in interacting pairs of galaxies than in isolated galaxies. These clues hint that activity in the nuclei of Seyfert galaxies is somehow triggered by collisions or interactions with other galaxies. You will find more such evidence as you study other kinds of active galaxies.

17-1b Double-Lobed Radio Sources

Beginning in the 1950s, radio astronomers found that some sources of radio energy in the sky consist of pairs of emitting regions. When optical telescopes studied the locations of these radio sources, they revealed galaxies located between the two radio source regions, and those galaxies were dubbed **double-lobed radio galaxies**. Unlike Seyfert galaxies, which produce intense radiation in their centers, these radio galaxies were emitting energy mostly from two external radio lobes.

Study **Cosmic Jets and Radio Lobes** and notice four important points and two new terms:

1. The shapes of radio lobes suggest that they are inflated by jets of excited gas emerging from the nucleus of the central galaxy. This is called the *double-exhaust model.* The presence of *hot spots* and synchrotron radiation shows that the jets are very powerful. (You encountered synchrotron radiation previously in Section 13-4e, in the context of supernova remnants.)
2. Active galaxies that have jets and radio lobes are often deformed or interacting with other galaxies.
3. The complex shapes of some jets and radio lobes can be explained by motions of the galactic nuclei that are producing them. A good example of this is 3C 31 (the 31st source in the



Panel a: NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/A. Evans (Univ. of Virginia)/NRAO/SUNY Stony Brook; Panel b: J. Mackenty and A. Stockton (IfA, Univ. of Hawai'i)

▲ **Figure 17-2** (a) The Seyfert galaxy NGC 7674 is part of a small, compact group of galaxies and is interacting with its smaller companion. Tidal tails are visible extending to the left and to the upper right. (b) The tidal tails are more easily visible in this enhanced image.

Statistical Evidence

How can statistics be useful if they can't be specific?

Some scientific evidence is statistical. Observations suggest, for example, that Seyfert galaxies are more likely to be interacting with a nearby companion than are normal galaxies. This is statistical evidence, so you can't be certain that any specific Seyfert galaxy will have a companion. How can scientists use statistical evidence to learn about nature when statistics contain built-in uncertainty?

Meteorologists use statistics to determine how frequently storms of a certain size are likely to occur. Small storms happen every year, but medium-size storms may happen on average only every ten years. Hundred-year storms are much more powerful but occur much less frequently—on average only once in a hundred years.

Those meteorological statistics can help you make informed decisions—as long as you understand the powers and limitations of statistics. Would you buy a house protected from a river by a levee that was not designed to withstand a hundred-year storm? In any one year, the chance of your house being destroyed would be only 1 in 100. You know the storm will hit eventually, but you don't know when. If you buy the house, a storm might destroy the levee the next year, but you might own the house for your whole life and never see a hundred-year storm. The statistics can't tell you anything about a specific year.

Before you buy that house, there is an important question you should ask the meteorologists. "How much data do you have about storms?" If they have only ten years of data, then they don't really know much about hundred-year storms. If they have three centuries of data, then their statistical data are significant.

Sometimes people dismiss important warnings by saying, "Oh, that's only statistics." Scientists can use statistical evidence if it passes two tests. It cannot be used to draw conclusions about specific cases, and it must be based on large enough data samples so the statistics are significant. With these restrictions, statistical evidence can be a powerful scientific technique.

Marko Georgiev/Getty Images

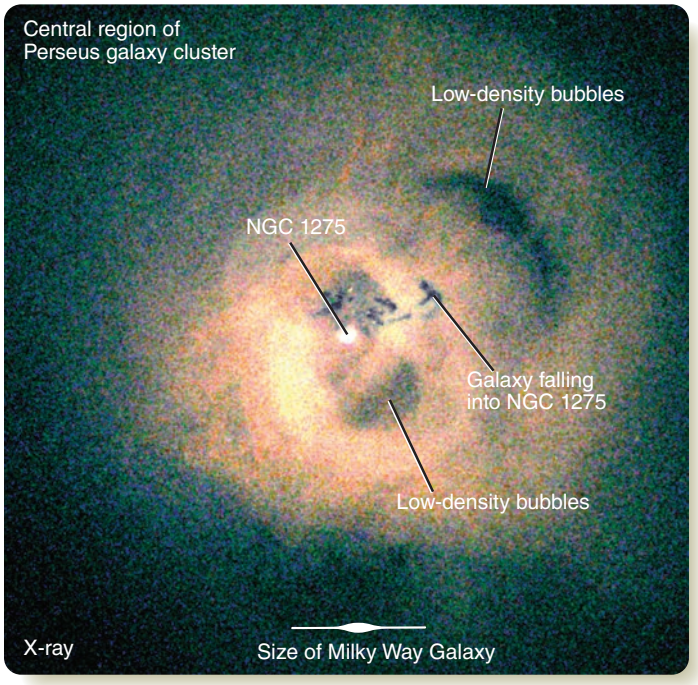
Statistics can tell you that a bad storm will eventually hit, but it can't tell you when.

Third Cambridge Catalog of Radio Sources) that has twisted radio lobes.

4 These jets are consistent with matter falling toward a central supermassive black hole and then somehow being ejected in two directions. In previous chapters you have seen similar jets produced by accretion disks around protostars, neutron stars, and stellar-mass black holes. Although the details are not entirely understood, the same process seems to be producing all of these jets.

The evidence shows that the nuclei of many galaxies are occupied by supermassive black holes with matter flowing inward. As the matter falls toward the center, it releases tremendous gravitational energy and forms hot accretion disks that can emit X-rays and eject jets in opposite directions.

Active galaxies can release huge amounts of energy. NGC 1275 (also called Perseus A) is one of the largest galaxies known, one of the thousands of galaxies that make up the Perseus Cluster. The nucleus of NGC 1275 is pumping out jets of high-energy particles, heating the huge gas cloud in which the galaxy cluster is embedded, and inflating low-density bubbles that distort the cloud (**Figure 17-3**). Galaxy NGC 1316 (also called Fornax A) has erupted to expel two lobes visible at radio wavelengths that would completely dwarf our Milky Way Galaxy (**Figure 17-4**). (The names "Perseus A" and "Fornax A" respectively signify the strongest radio



NASA/CXC/SAO/IoA/A. Fabian et al.

▲ **Figure 17-3** NGC 1275 in the Perseus Galaxy cluster is spewing out jets and streamers of high-energy particles that are inflating low-density cavities in the hot gas within the cluster. Proportionally, the entire galaxy cluster would be roughly 50 times larger in diameter than this image.

VLA/NRAO/AUI/NSF

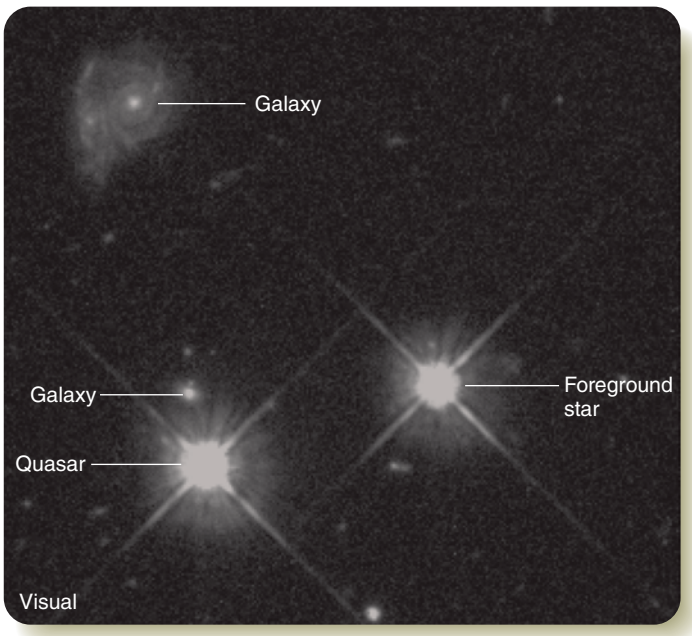
◀**Figure 17-4** Like cosmic fireworks, the galaxy NGC 1316 (Fornax A) has erupted to expel two lobes visible at radio wavelengths. The size of the Milky Way Galaxy is shown for comparison.

sources in those two constellations.) Observations indicate that the gas seen within galaxy clusters, referred to as the **intergalactic medium**, can be heated to temperatures of many millions of degrees by energy coming from AGN in the cluster.

17-1c Quasars

Through the 1950s, radio astronomers were studying celestial radio sources that were either supernova remnant gas clouds or distant radio galaxies, so they were surprised in the early 1960s when some radio sources turned out to look like stars in visual-wavelength photographs (**Figure 17-5**). First called *quasi-stellar objects,* they were soon referred to as **quasars**. Many more quasars have been found over the years, and most are radio silent. The radio astronomers stumbled over those emitting radio energy first because they were easy to notice. From the day they were discovered, quasars have been a challenge to understand.

The spectra of quasars were strange in that they contained a few unidentified emission lines superimposed on a continuous spectrum. In 1963, Maarten Schmidt at Caltech found that if a very large redshift was applied to the hydrogen Balmer lines, they could be made to agree with the lines in the spectrum of the bright quasar known as 3C 273.



NASA/ESA/STScI/AURA/NSF/C. Steidel (Caltech)

▲ **Figure 17-5** Quasars have starlike images clearly different from the images of even very distant galaxies. Although quasars look like stars, their spectra are unlike the spectra of stars or galaxies. The spikes on these images were produced by diffraction in the telescope.

KEY EQUATION 17-2 Cosmological Redshift

$$\text{redshift } z = \frac{\Delta\lambda}{\lambda_0}$$

(Note that this is similar to Eq. 7-3, the Doppler shift, although, as you will learn in the next chapter, the cosmological redshift is not a Doppler shift.)

Example: The Balmer H-beta line has a laboratory wavelength λ_0 of 486.1 nm. The line in the spectrum of 3C 273 that Schmidt guessed is H-beta has a wavelength of 562.9 nm. What is the redshift, z? ***Solution:*** The change in wavelength, $\Delta\lambda$, is the observed wavelength minus the laboratory wavelength:

$$\Delta\lambda = 562.9 - 486.1 = 76.8 \text{ nm}$$

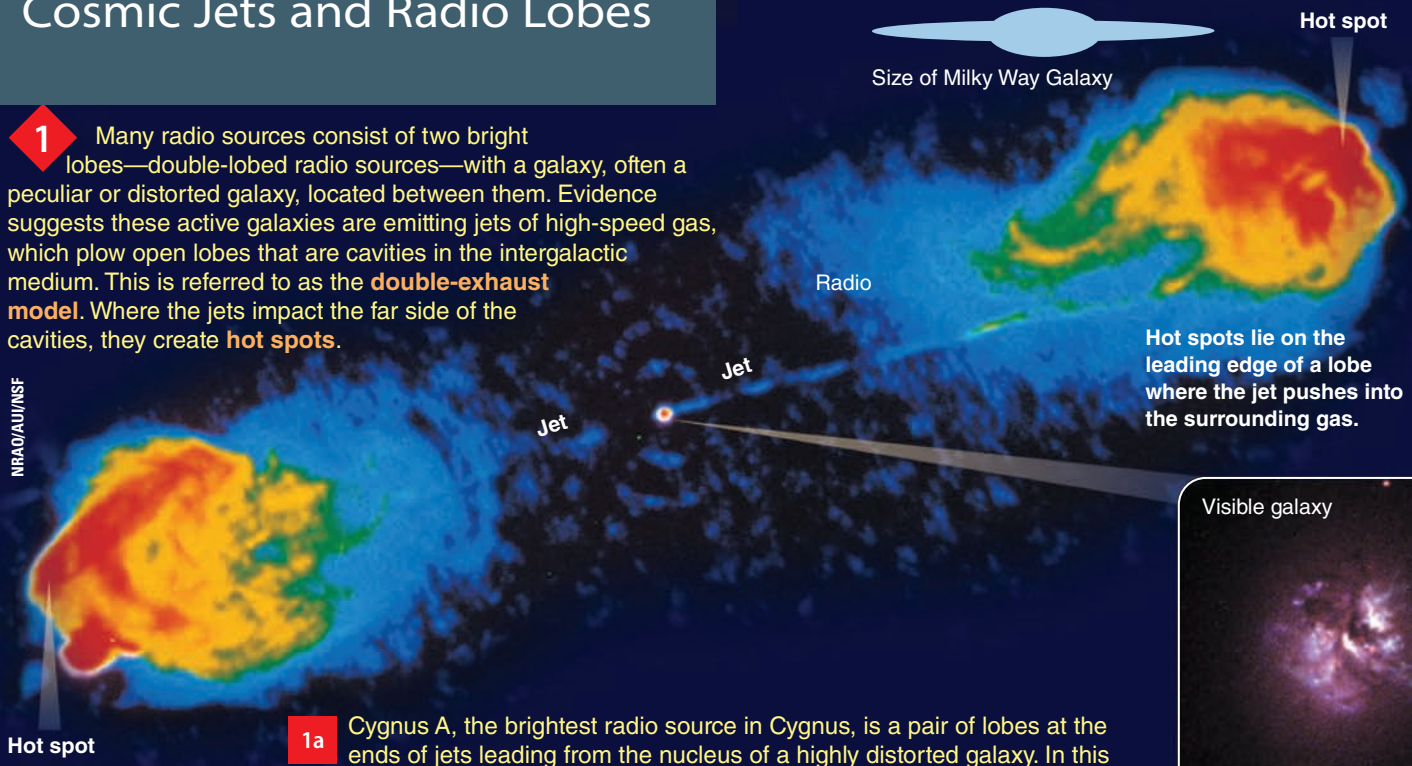
$$z = \frac{\Delta\lambda}{\lambda_0} = \frac{76.8}{486.1} = 0.158$$

At a redshift of 15.8 percent, Schmidt found that lines in the spectrum of 3C 273 clicked into place (**Figure 17-6**). Other quasar spectra quickly yielded to this approach, revealing even larger redshifts.

$$\text{redshift} = \frac{\Delta\lambda}{\lambda_0}$$

Cosmic Jets and Radio Lobes

1 Many radio sources consist of two bright lobes—double-lobed radio sources—with a galaxy, often a peculiar or distorted galaxy, located between them. Evidence suggests these active galaxies are emitting jets of high-speed gas, which plow open lobes that are cavities in the intergalactic medium. This is referred to as the **double-exhaust model**. Where the jets impact the far side of the cavities, they create **hot spots**.



NRAO/AUI/NSF

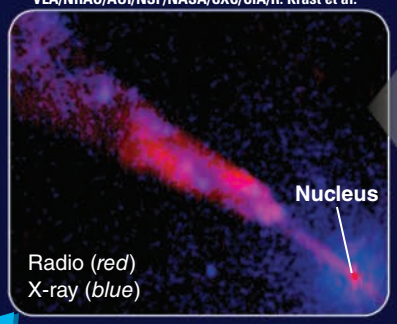
Hot spots lie on the leading edge of a lobe where the jet pushes into the surrounding gas.

Used with permission; R. A. E. Fosbury et al. 1998, "KNAW Colloquium on: The Most Distant Radio Galaxies."

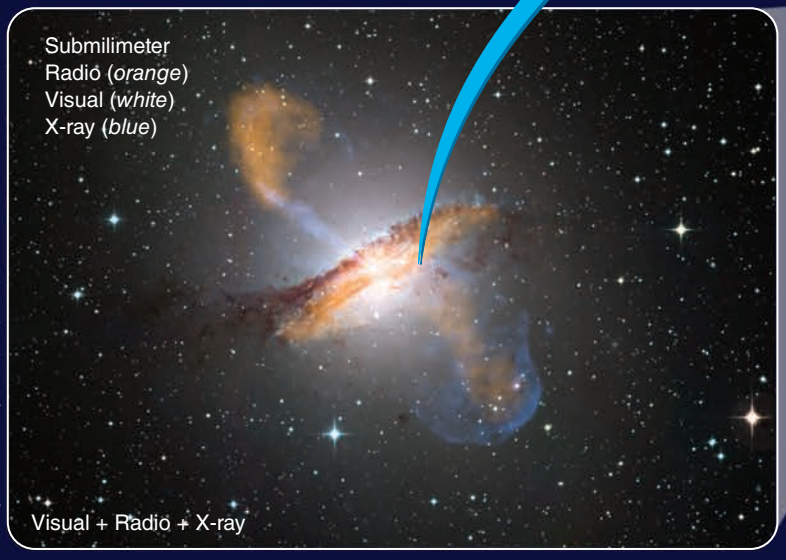
1a Cygnus A, the brightest radio source in Cygnus, is a pair of lobes at the ends of jets leading from the nucleus of a highly distorted galaxy. In this false-color image, the areas of strongest radio signals are shown in red and the weakest in blue. Because the radio energy detected is synchrotron radiation, astronomers conclude that the jets and lobes contain very-high-speed electrons, usually called *relativistic electrons*, spiraling through magnetic fields about 1000 times weaker than Earth's field. The total energy in a radio lobe is about 10^{53} J, which is what you would get if you turned the mass of a million Suns completely into energy.

2 The radio galaxy NGC 5128 lies between two radio lobes, and like many active galaxies, is strangely distorted. The dust ring rotates about an axis perpendicular to the ring, but the spherical cloud of stars rotates about an axis that lies in the plane of the ring. NGC 5128 appears to be two galaxies, a giant elliptical and a spiral, passing through each other. This has triggered multiple eruptions. A previous eruption evidently produced a large outer pair of lobes, and a more recent eruption produced an inner pair, closer to the center of the galaxy.

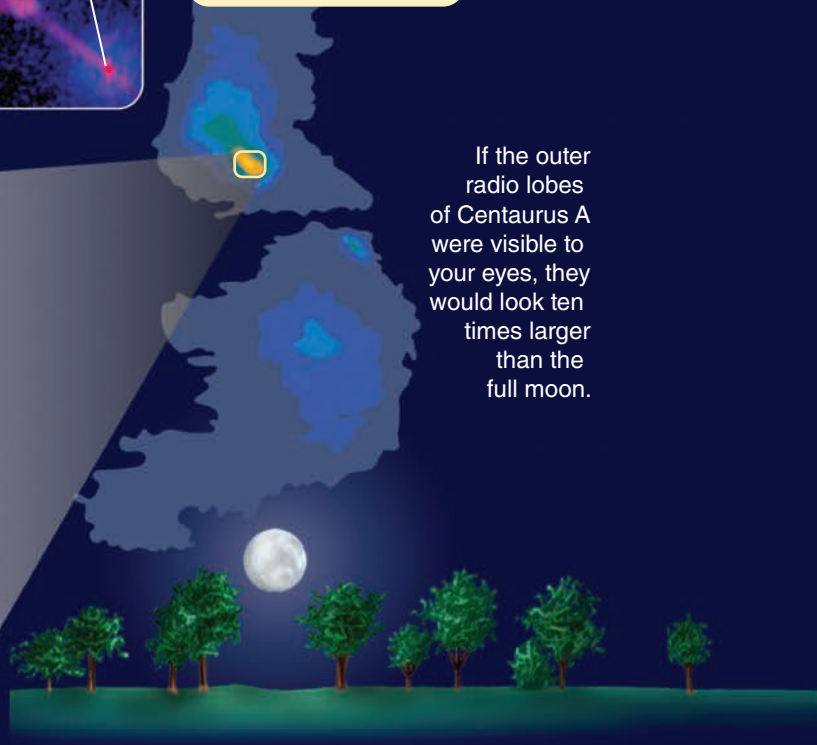
VLA/NRAO/AUI/NSF/NASA/CXC/CfA/R. Krast et al.



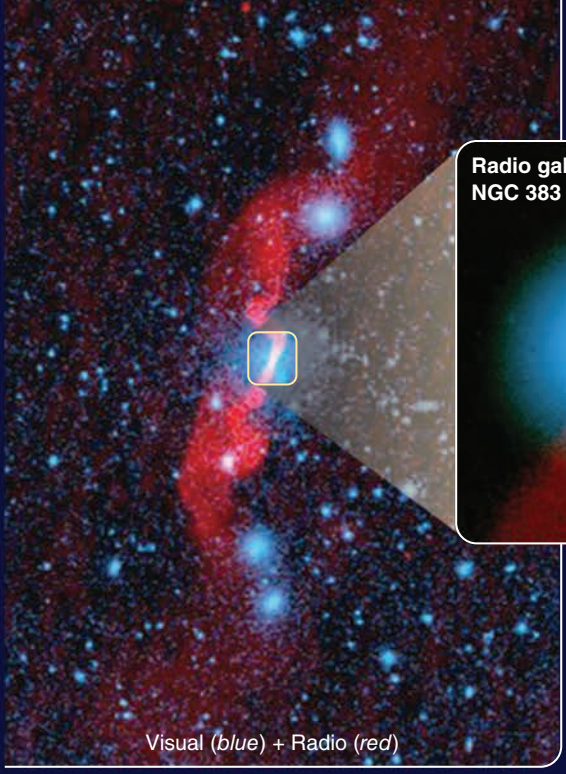
The combined radio and X-ray image at the left shows a high-energy jet at the very center of the galaxy pointing to the upper left into the northern radio lobe.



X-ray: NASA/CXC/CfA/R.Kraft et al.; Visual: ESO WFI; Radio (sub-millimeter): MPIfR/ESO/APEX/A. Weiss et al.



If the outer radio lobes of Centaurus A were visible to your eyes, they would look ten times larger than the full moon.



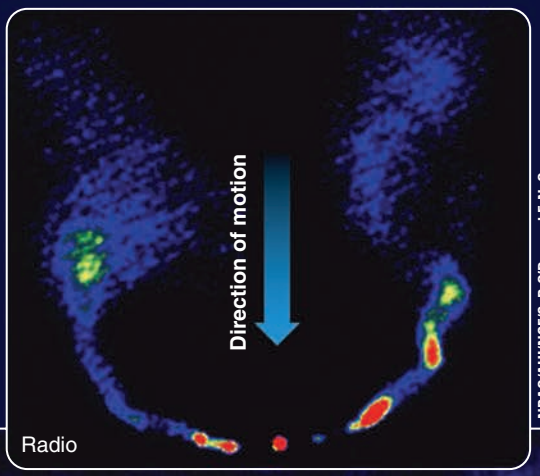
Radio: NRAO/AUI/NSF;
Visual: DSS

Radio: VLA/NRAO/AUI/NSF/A. Brindle & R. Laing;
Visual: NASA/ESA/STScI/AURA/NSF/Martel et al.

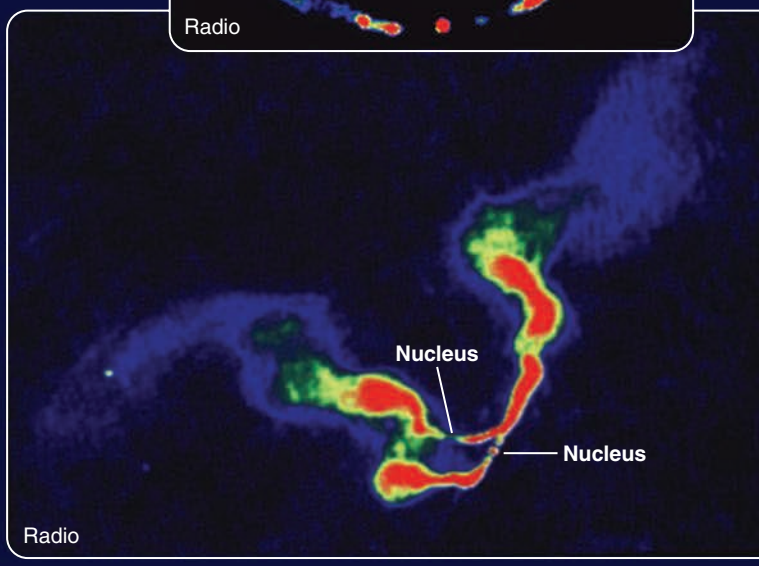
3a Radio galaxy 3C 31 is one of a chain of galaxies. It has ejected jets from its nucleus that twist, presumably because the active nucleus is orbiting another object such as the nucleus of a recently absorbed second galaxy.

3b The radio source 3C 75 is produced by two galaxies experiencing a close encounter. As the active nuclei orbit around each other, their jets twist and turn. This radio image only shows the jets and the nuclei; the host galaxies would be about the size of cherries at this scale.

3 The two radio jets from NGC 1265 are being left behind as the galaxy moves rapidly through the gas of the intergalactic medium. Twists in the jets' tails are understood to be caused by motions of the active nucleus from which the jets arise.



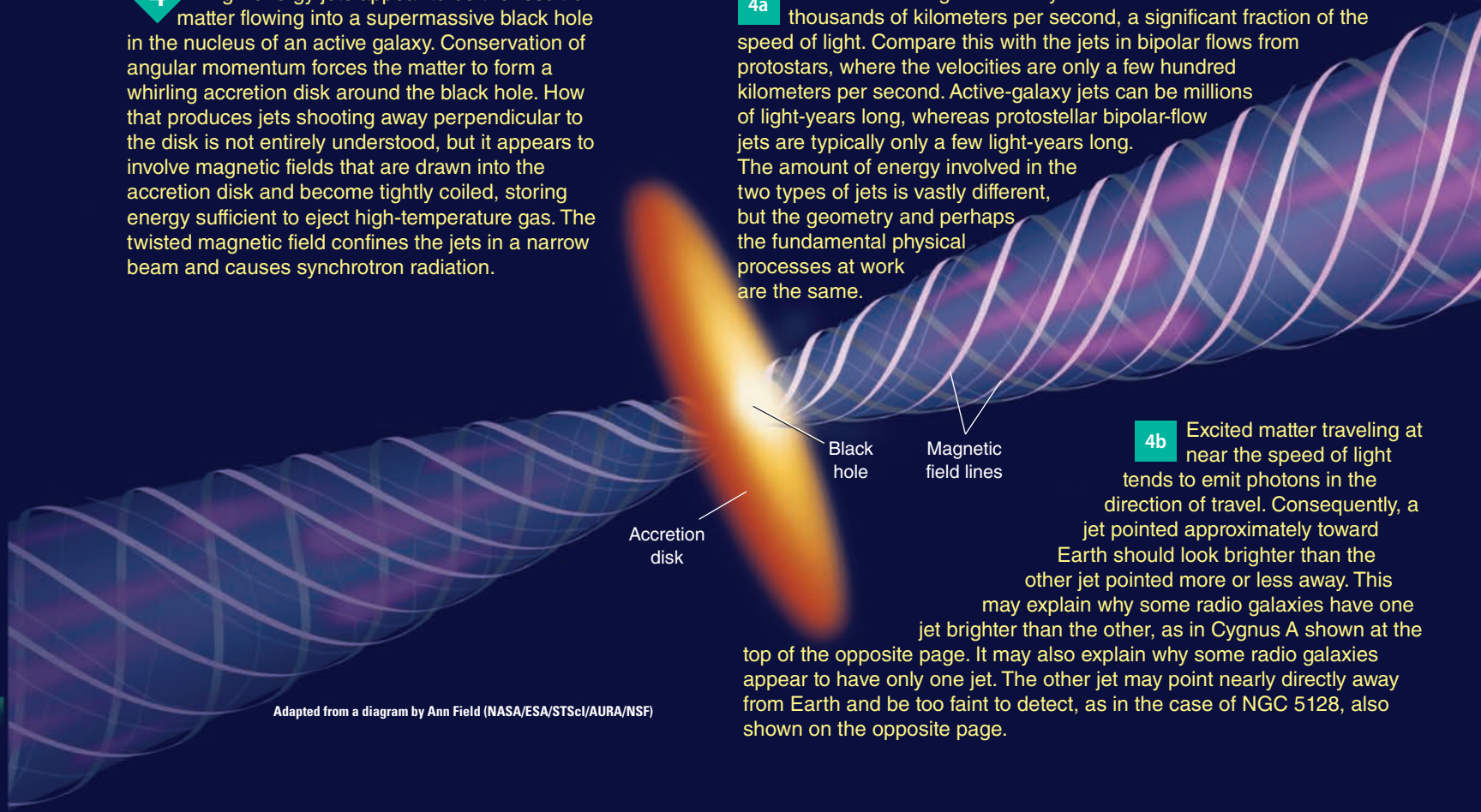
NRAO/AUI/NSF/C. P. O'Dea and F. N. Owen



NRAO/AUI/NSF/F. N. Owen, C. P. O'Dea, M. Inoue, and J. Eilek

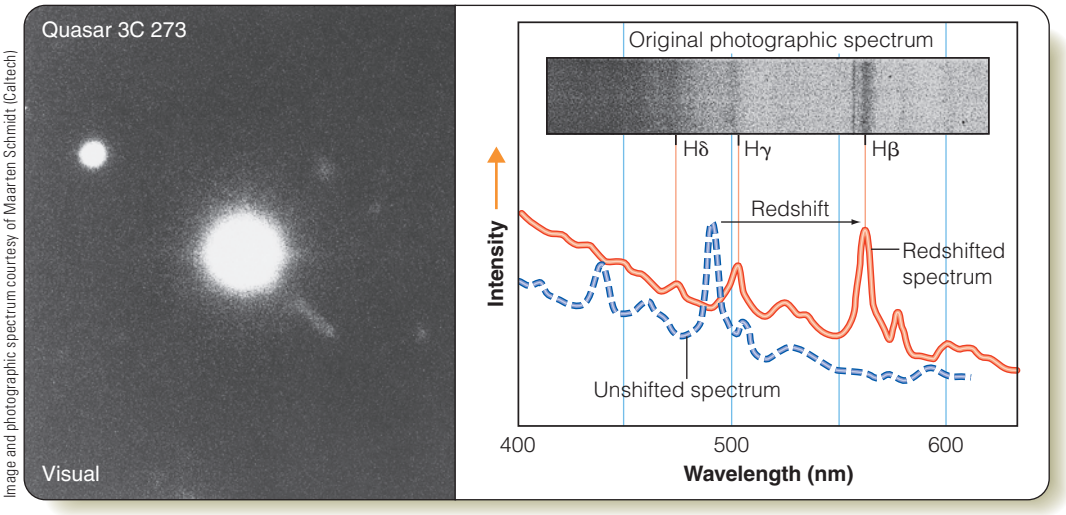
4 High-energy jets appear to be the result of matter flowing into a supermassive black hole in the nucleus of an active galaxy. Conservation of angular momentum forces the matter to form a whirling accretion disk around the black hole. How that produces jets shooting away perpendicular to the disk is not entirely understood, but it appears to involve magnetic fields that are drawn into the accretion disk and become tightly coiled, storing energy sufficient to eject high-temperature gas. The twisted magnetic field confines the jets in a narrow beam and causes synchrotron radiation.

4a Jets from active galaxies may have velocities from thousands to tens of thousands of kilometers per second, a significant fraction of the speed of light. Compare this with the jets in bipolar flows from protostars, where the velocities are only a few hundred kilometers per second. Active-galaxy jets can be millions of light-years long, whereas protostellar bipolar-flow jets are typically only a few light-years long. The amount of energy involved in the two types of jets is vastly different, but the geometry and perhaps the fundamental physical processes at work are the same.



Adapted from a diagram by Ann Field (NASA/ESA/STScI/AURA/NSF)

4b Excited matter traveling at near the speed of light tends to emit photons in the direction of travel. Consequently, a jet pointed approximately toward Earth should look brighter than the other jet pointed more or less away. This may explain why some radio galaxies have one jet brighter than the other, as in Cygnus A shown at the top of the opposite page. It may also explain why some radio galaxies appear to have only one jet. The other jet may point nearly directly away from Earth and be too faint to detect, as in the case of NGC 5128, also shown on the opposite page.



▲ **Figure 17-6** This image of 3C 273 shows the bright quasar at the center surrounded by faint fuzz. Note the jet protruding to the lower right. The original photographic plate containing a spectrum of 3C 273 with three hydrogen Balmer lines, H_δ, H_γ, and H_β. The spectrum is redshifted by 15.8 percent. The dashed line shows the position of the spectrum without a redshift.

The first quasars studied were the brighter ones, but surveys have found more, and more than a million are now known. Many quasar redshifts are greater than 1.0, and that may strike you as puzzling. The Doppler formula (Eq. 7-3) implies that such objects must have velocities greater than the speed of light, which is impossible. The answer is that the redshifts of galaxies and quasars are not produced by the Doppler effect. As you will discover in the next chapter, those redshifts are produced by the expansion of the Universe, and astronomers must use the equations of general relativity to interpret them. Redshifts greater than 1.0 are not a problem; they merely indicate great distance.

Although the quasars are far away, they are surprisingly bright. A typical galaxy at such a distance would be faint and extremely difficult to detect, but quasars show up on photographs as noticeable points of light. If you put the apparent brightness and huge distances of quasars into the magnitude–distance formula (Eq. 9-2), you will discover that quasars are incredibly luminous, with 10 to 1000 times the luminosity of a large galaxy like the Milky Way.

Soon after quasars were discovered, astronomers detected fluctuations in brightness over times as short as a few hours. Those rapid fluctuations show that quasars are small objects, no more than a few light-hours in diameter—smaller than our Solar System.

Astronomers trying to understand quasars faced a mystery: How could quasars be ultraluminous but also very small? What could make 10 to 1000 times more energy than a galaxy in a region as small as our Solar System? Since that time, new, large telescopes in space and on Earth's surface have revealed that quasars are often surrounded by hazy features with spectra resembling those of normal galaxies. Evidently, quasars are located in galaxies. In addition, radio telescopes have revealed that some quasars are ejecting jets and inflating radio lobes. The evidence is now overwhelming that quasars are exceptionally active nuclei of very distant galaxies. In other words, quasars are the most extreme kind of AGN.

Practicing **SCIENCE**

What evidence suggests that radio lobes are inflated by jets from the nucleus? To answer this question a scientist must combine seemingly unrelated clues about a system so huge that it cannot possibly be duplicated in a laboratory.

Of course, the strongest evidence relating radio lobes to jets is that, in many cases, a jet is actually detectable at visual or radio wavelengths extending from the center of a galaxy out into a radio lobe. But astronomers also note that hot spots occur on the outer edges of many radio lobes where a jet—whether it is visible or not—would collide with the intergalactic medium. In some cases where jets are detectable, they are curved and twisted by the orbital motion of the galaxy's active nucleus, which shows that they are being produced by gas expelled from the nucleus. Moreover, a rough "back-of-the-envelope" calculation shows that the amount of energy transported in the jets is about the right amount to produce the radio lobes.

Evidence is the key to understanding science. Now, use evidence to address another basic question about the nature of AGN and quasars: ***What evidence shows that AGN and quasars must be small?***

17-2 Supermassive Black Holes

In previous chapters, you learned that many galaxies, including our own Milky Way Galaxy, contain supermassive black holes at their centers. Why do such objects produce eruptions, and how did they form?

17-2a Disks and Jets

Matter flowing inward toward a black hole spins very fast and becomes very hot. Whatever amount of spin the matter has to begin with is amplified by conservation of angular momentum as it approaches the black hole. As is true for stellar-mass black holes, the incoming matter forms a flattened accretion disk around the central compact object. Supermassive black holes have stronger gravity than stellar-mass black holes, producing faster spins for the infalling matter and higher temperatures. Even a supermassive black hole is surprisingly small. A 10-million-solar-mass black hole would be only one-fifth the diameter of Earth's orbit, so the matter can get close to the black hole and orbit very fast. The infalling matter heats up because it picks up speed, and as it collides with other matter, those high velocities become thermal energy. That is, matter falling inward converts gravitational energy into thermal energy and becomes hot. This is the same process you learned about in Section 11-1c that heats a collapsing interstellar cloud to produce a protostar.

Theoretical calculations predict that that the high temperature "puffs up" the inner part of the disk and makes it thick. Closer to the black hole, an orbiting particle is unstable and must spiral into the black hole, so the black hole is hidden deep inside an empty cavity at the center of the accretion disk. Model calculations indicate that the disk is thinner and cooler at intermediate distances from the black hole, but the outermost part of the disk is a fat, cool torus (doughnut shape) of dusty gas.

Astronomers can't see black holes, but in some active galaxies the *Hubble Space Telescope* can detect the outer parts of the central disks (**Figure 17-7**). Spectra reveal the speed of rotation, and Kepler's third law yields the mass of the central object. Some supermassive black holes have masses of a few million solar masses, like the one in the center of our Milky Way Galaxy, but the most massive contain billions of solar masses.

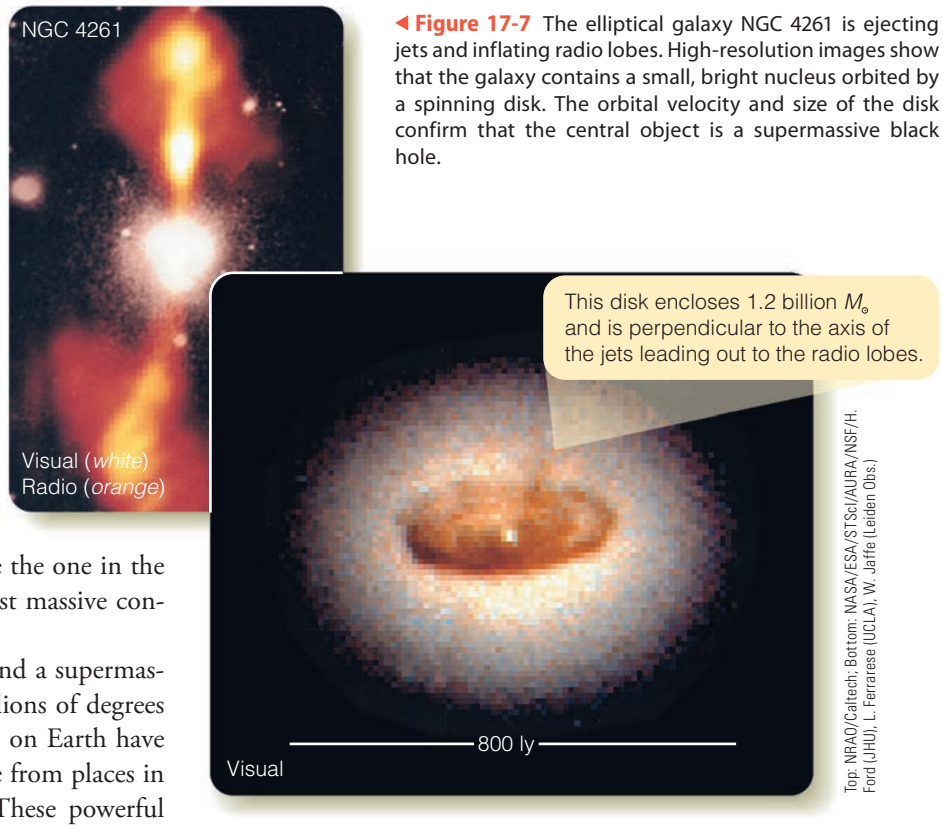
The inner part of the accretion disk around a supermassive black hole can reach temperatures of millions of degrees and emit X-rays. In fact, cosmic ray detectors on Earth have observed ultra-high-energy particles that come from places in the sky occupied by active galactic nuclei. These powerful supermassive black holes and their hot accretion disks may be spraying the Universe with high-speed particles, each packing the wallop of a major league baseball pitch.

As is the case for stellar-mass black holes, the mechanism whereby a supermassive black hole and its accretion disk produce jets of gas and radiation is not well understood, but magnetic fields seem to be involved. Because the disk is at least partially ionized, magnetic fields are trapped in the gas of the disk, drawn inward, and wound up. Theorists suggest that this creates powerful magnetic tubes extending along the axis of rotation, perpendicular to the accretion disk, channeling hot gas and radiation outward in opposite directions. The jets seem to originate close to the supermassive black hole and are then focused and confined by the enclosing magnetic tubes.

The mechanism that produces jets is understood in only a general way, but astronomers are now trying to work out the details. How can supermassive black holes explain all of the different kinds of active galaxies that are observed?

17-2b A Unified Model of Active Galaxies

When a field of research is young, scientists find many seemingly different phenomena, such as Seyfert galaxies, double-lobed radio galaxies, quasars, and cosmic jets. As the research matures, scientists begin seeing similarities and eventually are able to unify the different phenomena as different aspects of a single process. This organization of evidence and theory into logical arguments that explain how nature works is the real goal


NGC 4261
Visual (*white*)
Radio (*orange*)
This disk encloses 1.2 billion $M_\odot$ and is perpendicular to the axis of the jets leading out to the radio lobes.
800 ly
Visual
Top: NRAO/Caltech; Bottom: NASA/ESA/STScI/AURA/NSF/H. Ford (JHU), L. Ferrarese (UCLA), W. Jaffe (Leiden Obs.)

◀ **Figure 17-7** The elliptical galaxy NGC 4261 is ejecting jets and inflating radio lobes. High-resolution images show that the galaxy contains a small, bright nucleus orbited by a spinning disk. The orbital velocity and size of the disk confirm that the central object is a supermassive black hole.

of science. Astronomers studying active galaxies have developed a **unified model** of AGN that is well supported by evidence. A supermassive black hole is the centerpiece.

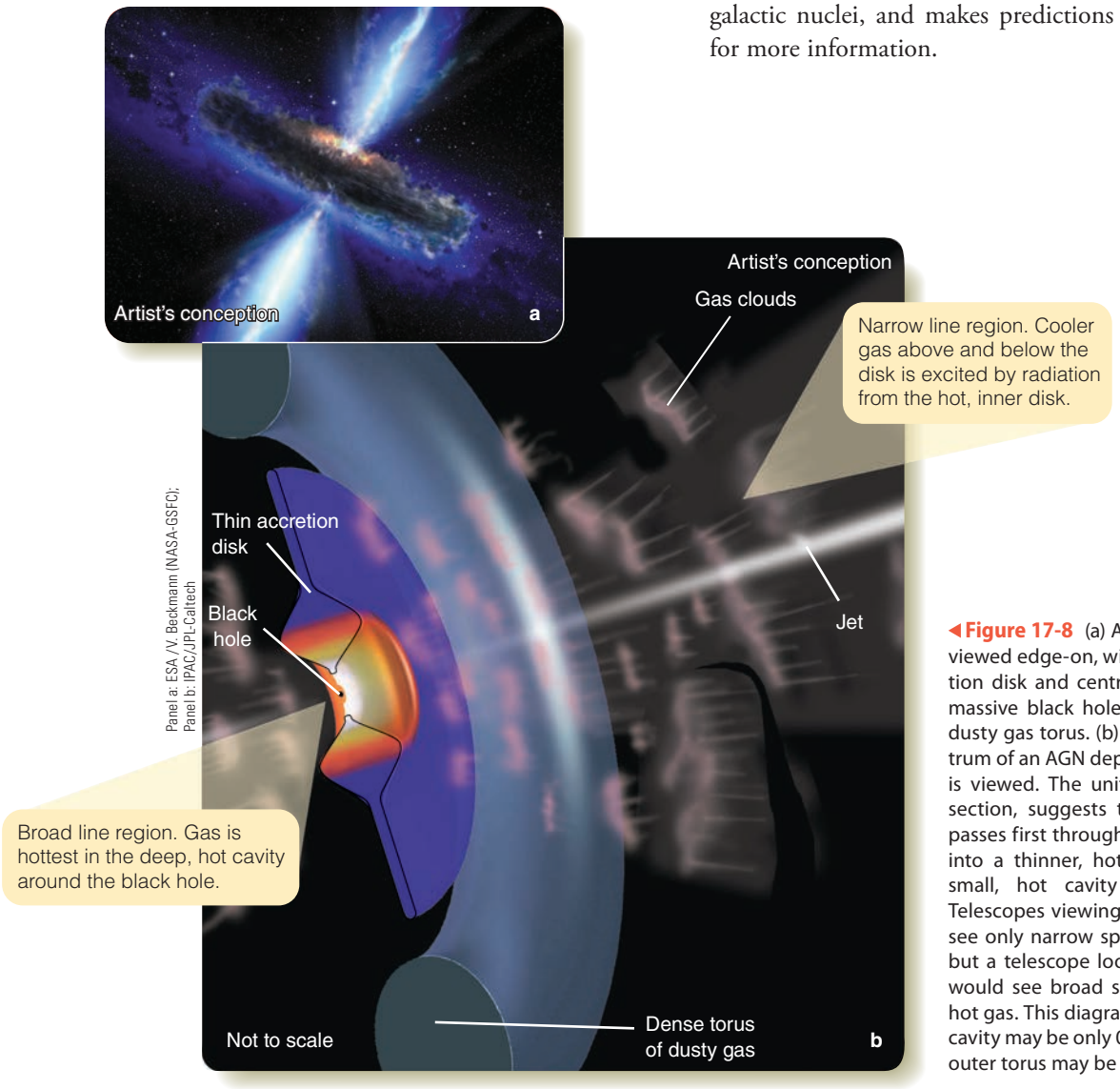
According to the unified model, what you see when you view the nucleus of an active galaxy depends on how the black hole's accretion disk is tipped with respect to your line of sight. You should note that the accretion disk may be tipped at a steep angle to the plane of its galaxy, so just because you see a galaxy face-on doesn't mean you are looking at the accretion disk face-on.

If you view the accretion disk edge-on, you cannot see the central area at all because the outermost thick dusty torus blocks your view. Instead, you see gas lying above and below the central disk, gas that is excited by intense radiation coming from the black hole inside the central cavity. Because this gas is relatively far from the center, it is cooler, orbits more slowly, has smaller Doppler shifts, and thus produces narrower spectral lines (**Figure 17-8**). This is the explanation for type 2 Seyfert galaxies.

If the accretion disk is tipped slightly to the line of sight, you may be able to see "over" the outer torus into the central cavity to observe some of the intensely hot gas there. This region emits broad spectral lines because the hot gas is orbiting at high velocities near the central mass and the high Doppler shifts smear out the lines. This is understood to be the explanation for type 1 Seyfert galaxies.

What happens if you look directly, face-on, into the central cavity? According to the unified model, you would be looking directly "down the dragon's throat" into one of the two high-energy jets blasting away perpendicular to the accretion disk. A few such objects have been found; they have featureless spectra and fluctuate in brightness in only hours.

The unified model is far from complete. The actual structure of accretion disks is poorly understood, as is the process by which the disks produce jets. The unified model does not explain all of the differences among active galaxies and quasars. Rather, it provides a logical framework that organizes the observations already made, provides some clues to what is happening in active galactic nuclei, and makes predictions that encourage searches for more information.


Artist's conception
a
Artist's conception
Gas clouds
Narrow line region. Cooler gas above and below the disk is excited by radiation from the hot, inner disk.
Thin accretion disk
Black hole
Jet
Broad line region. Gas is hottest in the deep, hot cavity around the black hole.
Not to scale
Dense torus of dusty gas
b
Panel a: ESA /V. Beckmann (NASA-GSFC); Panel b: IPAC/JPL-Caltech

◀ **Figure 17-8** (a) Artist's conception of an AGN viewed edge-on, with the view of the hot accretion disk and central cavity around the supermassive black hole blocked by the outer cool dusty gas torus. (b) Features visible in the spectrum of an AGN depend on the angle at which it is viewed. The unified model, shown in cross section, suggests that matter flowing inward passes first through a large, opaque torus; then into a thinner, hotter disk; and finally into a small, hot cavity around the black hole. Telescopes viewing such a disk edge-on would see only narrow spectral lines from cooler gas, but a telescope looking into the central cavity would see broad spectral lines formed by the hot gas. This diagram is not to scale. The central cavity may be only 0.01 pc in radius, whereas the outer torus may be 1000 pc in radius.

17-2c Triggering Eruptions

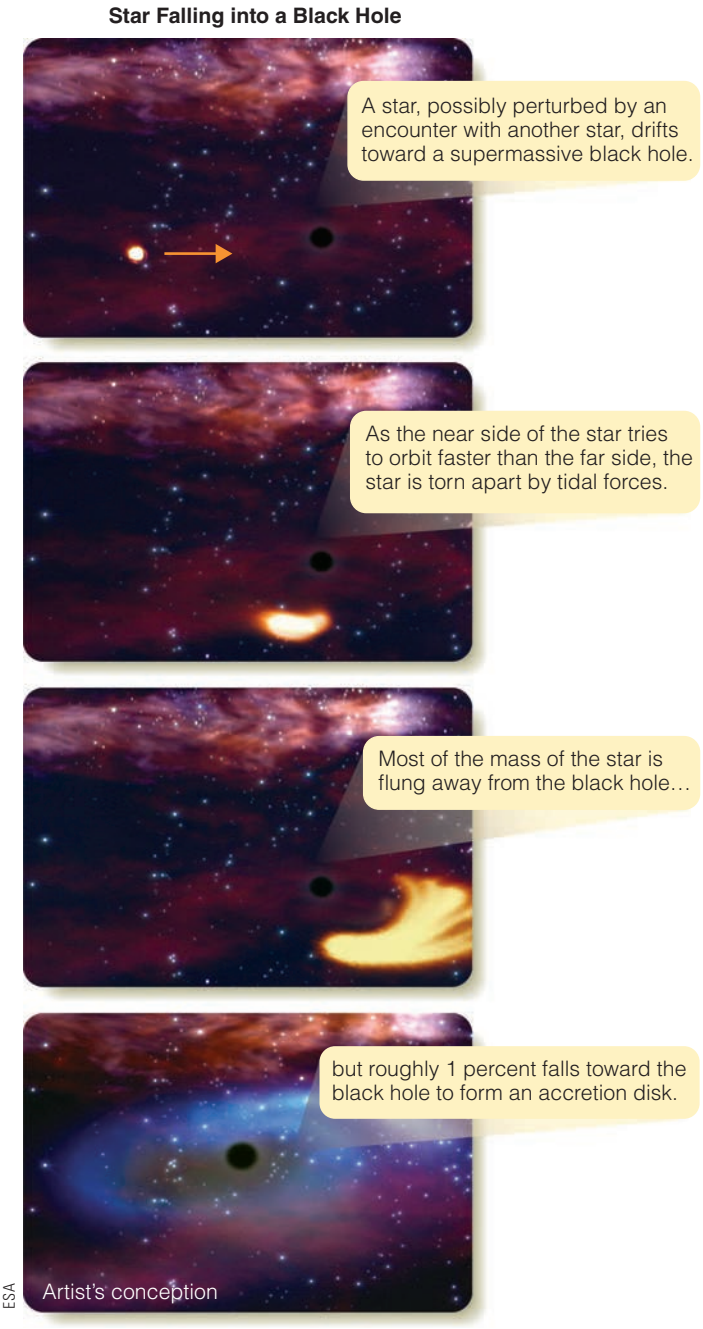
As you have already learned, most large galaxies contain supermassive black holes at their centers, but only a small percentage of galaxies have active galactic nuclei. Evidently, most supermassive black holes in the centers of galaxies are dormant. Why is that? And what occasionally wakes some of them up?

Recall from Section 14-2 that real black holes are not like cartoon black holes; they do not aggressively swallow everything around them. Actually, a black hole with no matter falling into it is rather difficult to find, only betraying its presence by making nearby objects move more rapidly than they otherwise would. So, what would suddenly cause a black hole to flare up? The answer is something that you studied back in Section 5-2f—tides. In the previous chapter, you saw how tides twist interacting galaxies and rip matter away into tidal tails. Mathematical models show that those same interactions can also throw matter inward. A sudden flood of matter flowing inward and producing an accretion disk around a supermassive black hole would trigger an eruption. This explains why many active galaxies are distorted; they have been twisted by tidal forces as they interacted or merged with another galaxy. Some active galaxies have nearby companions, and you can suspect that the companions are guilty of tidally distorting the other galaxy and triggering an eruption by tossing material toward its central supermassive black hole.

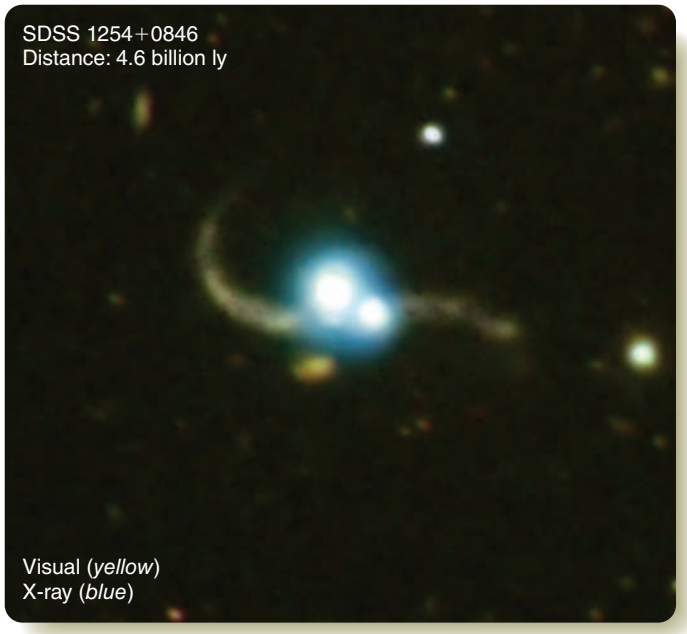
Tidal interactions between galaxies span distances of 100,000 ly or more, but, as you learned in Section 14-2c, the same bit of physics becomes important when matter comes very close to a black hole. **Figure 17-9** shows how a passing star could be ripped apart by the tidal forces near a supermassive black hole, and at least partially consumed. Inflowing gas, dust, and an occasional star would be an energy feast for a supermassive black hole.

Observations support the idea that galaxy interactions and mergers trigger active galactic nuclei. Earlier in this chapter you learned that many active galaxies have distorted shapes. And, although quasars are generally so far away that images of their host galaxies are difficult to obtain, quasars, like AGN nearer to Earth, are found to be embedded in galaxies that are distorted and are near other distorted galaxies. Furthermore, regarding galaxy mergers, more than one-third of active galaxy nuclei in one survey were found to involve double objects. That is, the active nucleus is orbiting an object of similar mass. Apparently, these objects are the result of mergers in which the original supermassive black holes of the two galaxies have sunk to the center of a new combined galaxy and are presumably spiraling toward each other as they emit gravity waves (**Figure 17-10**).

Galaxy IC 2497 provides a dramatic example of galaxy interactions and active nuclei. Near the galaxy lies a glowing, green cloud of gas. Named Hanny's Voorwerp, the object is part of a streamer of gas, dust, and stars pulled out of the galaxy by a recent interaction. Where a beam of energy coming from the active nucleus of the galaxy shines on the streamer, the gas is excited to glow green in forbidden emission lines (**Figure 17-11**). Presumably the AGN was triggered into activity by the same tidal interaction that pulled away the streamer.


Star Falling into a Black Hole
A star, possibly perturbed by an encounter with another star, drifts toward a supermassive black hole.
As the near side of the star tries to orbit faster than the far side, the star is torn apart by tidal forces.
Most of the mass of the star is flung away from the black hole…
but roughly 1 percent falls toward the black hole to form an accretion disk.
Artist's conception
ESA

▲ **Figure 17-9** Orbiting X-ray telescopes observing active galaxies sometimes detect X-ray flares equaling the energy of a supernova explosion. Such flares are evidently caused when a star wanders too close to the supermassive black hole at the center of the galaxy and tidal forces rip the star apart.



X-ray: NASA/CXC/CfA/P. Green et al.; Visual: Carnegie Obs./Magellan/J. Mulchaey et al.

▲ **Figure 17-10** This X-ray image shows two quasars (*center*) embedded in two galaxies in the act of merging. Tidal forces between the galaxies have flung out two tidal tails shown in the visual-wavelength image. The tidal tails are a sure sign of galaxies merging, and that has triggered the two supermassive black holes into eruption as quasars.

NASA/ESA/STScI/AURA/NSF/W. Keel (Univ. of Alabama), Galaxy Zoo Team

17-3 A History of Galaxies and Supermassive Black Holes

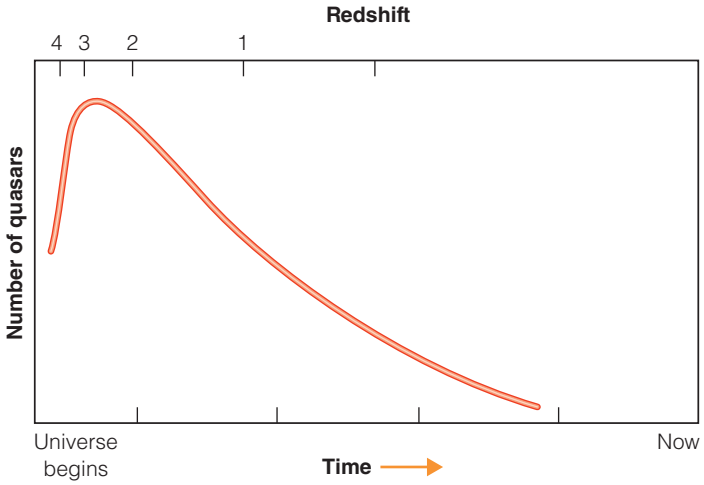
When you look at a photo of galaxies with large redshifts, the look-back time is large, and you see the Universe as it was long ago. The light journeying from such a great distance carries information about how the galaxies formed and developed.

In the next chapter, you will see evidence that the Universe began 13.8 billion years ago and that it has been expanding ever since. Model calculations indicate that, within a few hundred million years after the Universe began, the first protogalaxy gas clouds started forming stars. Some of those protogalaxies would then have combined to produce the great star clouds that became the central bulges of galaxies. Models also suggest that some of the first gas clouds would have contracted to form supermassive black holes. In fact, some evidence suggests that the black holes formed first and pulled matter inward to help form central bulges.

As you have learned, quasars and other types of active galaxies are triggered into eruption by tidal interactions, collisions, and mergers of galaxies, processes that throw matter into central supermassive black holes. Therefore, you might expect that, when the Universe was young and had not yet expanded very much, galaxies were closer together and would have interacted more often. And, from your understanding of how the Milky Way and other galaxies formed (look back to Sections 15-5e and 16-3c), you might hypothesize that small galaxies falling into, combining with, and forming halos around larger galaxies would have fed matter inward to the supermassive black holes. Matter flooding into those black holes would have triggered powerful outbursts, which would be visible from Earth today at large look-back times. But what do the observations say?

Quasars are most common with redshifts over 2 and less common with redshifts above 3 (**Figure 17-12**). The most distant quasars found so far have redshifts greater than 7, but such high-redshift quasars are quite rare. Evidently, at redshifts between 2 and 3, galaxies were actively growing, colliding, and merging, and quasars were about 1000 times more common than they are now. These observations imply that in the first half of the history

◀ **Figure 17-11** A beam of energy from active galaxy IC 2497 (*top*) illuminates part of an infalling streamer of matter, exciting the gas to glow green. The glowing cloud was discovered by Dutch high school teacher Hanny van Arkel while using her personal computer with the Galaxy Zoo project (www.galaxyzoo.org/), which recruits amateurs to classify the millions of galaxies found in deep surveys. The object's name, Hanny's Voorwerp, means "Hanny's Object" in Dutch. Now classified as a quasar, the active nucleus may have turned on as recently as 200,000 years ago.



▲ **Figure 17-12** At large look-back times, you see the Universe as it was long ago. Redshifts between about 2 and 3 look back to a time when galaxies were closer together and were actively forming, merging, and interacting. That produced quasars. More recently, quasars have become less common.

of the Universe, each galaxy merged with other galaxies as often as three times every billion years. Most mergers were with smaller galaxies, but a few would have been with larger galaxies, triggering the biggest eruptions detectable from Earth today as quasars. Nevertheless, even during that "age of quasars," quasar eruptions must have been unusual. At any one time, only a small fraction of galaxies had quasars erupting in their nuclei. As smaller galaxies were gobbled up and the expansion of the Universe carried the galaxies away from each other, galaxy formation became less active, interactions between galaxies became rare, and quasars became even less common.

Astronomers hypothesize that few quasars are seen with redshifts greater than 3 because that is looking back to an age when the Universe was so young it had not yet formed many galaxies. Some of the most distant quasars could have been caused by the initial formation of a few supermassive black holes before galaxies formed, but those extremely distant objects are difficult to study with existing telescopes. This hypothesis will have to be checked using larger telescopes expected to be built in the near future (look back to Sections 6-3 and 6-4).

Astronomers have been able to measure the masses of a few dozen supermassive black holes by observing the speed of matter whirling around them. Interestingly, as was mentioned in the previous chapter, those masses are correlated with the masses of the host galaxies' central bulges. In each case, the mass of the black hole has about 0.5 percent of the mass of the surrounding central bulge for disk galaxies, or about 0.5 percent of the mass of the galaxy for elliptical galaxies. But, in the case of disk galaxies, there is no relationship between the masses of the black holes and the masses of the disks. You can see that this means there is also no relationship for disk galaxies between the mass of the central bulge and the mass of the disk. These statistics provide intriguing clues into how galaxies formed. The formation process for supermassive black holes in the centers of galaxies seems to have been directly connected with the formation of their central bulges but not to the formation of their disks. What does this mean?

Astronomers can get some insight into the distant era of galaxy formation by observing current activity in places like the Perseus Cluster (Figure 17-4), the brightest galaxy cluster X-ray source in the entire sky. (The Perseus Cluster is only 240 million ly from Earth, which counts as the present day relative to the age of the Universe, 13.8 billion years.) In the center of the Perseus Cluster, giant elliptical galaxy NGC 1275's supermassive black hole is undergoing an exceptionally powerful eruption. That eruption has heated the cluster's intergalactic medium to such a high temperature that it is escaping from NGC 1275. As a result, the supply of gas falling into the central supermassive black hole will be interrupted soon, astronomically speaking. In this way, a supermassive black hole can choke off an accretion flow and curtail its own growth until the surrounding gas can cool and begin to fall in again.

Astronomers speculate that there may be a maximum growth rate for supermassive black holes set by this cycle of self-limiting activity. Back in the quasar era, when galaxies were first being assembled, the formation of central bulges and supermassive black holes evidently would have been a violent process. The increasing power of outbursts from the growing central black hole in a forming galaxy eventually would have been able to push away infalling gas and limit the material available to form the stars of the central bulge, halting formation of the bulge. This may be the explanation for the relation between the masses of supermassive black holes and the masses of the central bulges of the galaxies in which they reside.

In a spiral galaxy such as the Milky Way, the disk with its spiral arms and bright stars seems a dominant part of the galaxy. But recall from Section 15-5 that the disk formed late as matter gradually settled into the galaxy. By that stage, the central bulge and supermassive black hole had already formed, and the gradual development of the outer disk probably didn't involve violent, self-limiting eruptions. That would be consistent with the observation that the masses of spiral galaxy disks are not proportional to the masses of their central bulges and supermassive black holes.

"Then where are all the dead quasars?" an astronomer asked recently. There is no way to get rid of supermassive black holes, so all of the galaxies that had short-lived quasars must still have those supermassive black holes. Where are all those dead quasars today? You know where to look—the centers of ordinary, nonactive galaxies.

The majority of large galaxies probably contain supermassive black holes at their centers, but the black holes have consumed most of the nearby gas, dust, and stars and are now

dormant. Our own Milky Way Galaxy could have had a quasar at its nucleus long ago, but now its central supermassive black hole is on a strict diet. A slow trickle of matter flowing into the black hole could explain the mild activity seen there. Occasional interactions between galaxies can throw matter inward and nudge the dragon, awaking a sleeping supermassive black hole into eruption as a Seyfert galaxy or a double-lobed radio galaxy.

The active galaxies are not a rare kind of galaxy. They are normal galaxies passing through a short-duration stage that many galaxies experience: They are not really peculiar. They are an important part of the story of the formation and evolution of galaxies.

Practicing SCIENCE

What do the distorted shapes of active galaxies mean?
This question leads to a more fundamental question: Why are some galaxies with central supermassive black holes active, although most are not active?

Seyfert galaxies and the galaxies located between double radio lobes are often distorted and have tidal tails. Also, the galaxies that are faintly visible around quasars are also typically twisted out of shape. These distortions must be caused by tidal interactions with nearby galaxies or by mergers, and such tides can throw matter inward toward the supermassive black hole at the center of a galaxy and trigger an eruption.

Now ponder a related question aimed at understanding the ultimate cause of active galaxies and quasars: ***Why were quasars relatively common when the Universe was young, and why are they rare today?***

WHAT ARE WE?

Changed

Next time you are in a shopping mall, glance at the people around you. How many of them, do you suppose, know, or would care, that galaxies can erupt, that there was an age of quasars, that our Milky Way Galaxy has a sleeping quasar in its nucleus, that "our" quasar was once active and affected the growth and evolution of the galaxy in which we live?

The vast majority of people have no idea how their own lives fit into the story of the Universe. Most people don't know what they are. They eat pizza and watch TV without understanding that they are part of a Universe in which supermassive black holes form in the centers of galaxies and occasionally erupt in titanic explosions.

Astronomy is changing you. As you learn more about stars and galaxies and quasars, you are learning more about yourself and your connection with nature. *Perspective* can mean a view of things in their true relationships. As you study astronomy, you are gaining perspective. Our Galaxy, our Sun, our planet, and the local shopping mall take on new meaning when you think astronomically.

Study and Review

SUMMARY

17-1 Active Galactic Nuclei

- First called **radio galaxies**, **active galaxies** are now known to emit energy at many wavelengths. Because the energy comes from their nuclei, they are said to contain **active galactic nuclei (AGN)**.
- **Seyfert galaxies** are one type of AGN. They are spiral galaxies with small, highly luminous nuclei, and their spectra tell us that those nuclei contain highly excited gas.
- **Double-lobed radio galaxies** are another type of AGN. A double-lobed radio galaxy emits radio energy from lobes, which are located on either side of the galaxy. As discussed in the **double-exhaust model**, the two lobes are inflated by two jets, which are ejected from opposite sides of the galaxy's nucleus. **Hot spots** form where the two jets run into the surrounding gas of the **intergalactic medium**.
- **Quasars** are the brightest of AGN. They have a starlike appearance, and because their spectra reveal large redshifts they must be distant galaxies. To be visible at such distances, quasars must be ultraluminous. Because quasars have rapid fluctuations in brightness, they must be small. A typical quasar has a diameter of only a few light-hours, comparable in size to our Solar System.

17-2 Supermassive Black Holes

- Matter flowing into a supermassive black hole must conserve angular momentum and form an accretion disk. The spinning accretion disk pulls in and winds up the magnetic field. By a process not yet fully understood, the magnetic field ejects two jets of high-speed gas and radiation in opposite directions from the disk. The magnetic field also focuses the jets along the axis of rotation.

- According to the **unified model**, what you see depends on the tilt of the accretion disk relative to your line of sight. If you see into the narrow cavity at the center, then you see broad spectral lines produced by hot, high-velocity gas.
  If you see the disk more edge-on and you cannot see into the cavity, then you see only narrow spectral lines. If the jet from the black hole points directly at you, then you see a strong but featureless spectrum that flickers in brightness over just a few hours.
- Supermassive black holes erupt only when large amounts of matter flow inward. Therefore, most galaxies have dormant nuclei as only a few small gas blobs fall inward occasionally. During galaxy interactions and collisions, tides can throw matter into the galaxy centers and trigger eruptions. This explains why active galaxies are often distorted or have nearby companions.

17-3 A History of Galaxies and Supermassive Black Holes

- Because the masses of supermassive black holes are correlated with the masses of the central bulges in which they reside, astronomers conclude that supermassive black holes formed when the first gas clouds fell together to form the central bulges, soon after the Universe began.

REVIEW QUESTIONS

1. How are active galaxies and radio galaxies related?
2. Are active galaxies rare? Are they peculiar? If so, explain. If not, why not?
3. Do all active galaxies have active nuclei?
4. Are all AGNs active galaxies? Are all active galaxies AGNs?
5. What statistical evidence suggests that Seyfert galaxies have suffered recent interactions with other galaxies?
6. Compare and contrast double-lobed radio galaxies versus Seyfert galaxies.
7. What evidence shows that the energy source in a double-lobed radio galaxy lies at the center of the galaxy?
8. How does a galaxy nucleus become active?
9. What are the observational characteristics of active galaxies?
10. How does the peculiar rotation of NGC 5128 help explain the origin of this active galaxy?
11. What evidence shows that quasars are ultraluminous and small?
12. What evidence indicates that quasars occur in distant galaxies?
13. According to the text, 3C 273 has a 15.8 percent redshift. Is that object near or distant? How do you know?
14. What causes the redshifts we see in spectra of quasars?
15. What makes a supermassive black hole '"active"?
16. How does the unified model explain the two kinds of Seyfert galaxies?
17. Why are few quasars found at low redshifts and at high redshifts, although many are found at intermediate redshifts of approximately 2 to 3?
18. Are quasars common near the Milky Way Galaxy? Why or why not?
19. Why did galaxies collide more often in the distant past than they do now?
20. If you see a featureless spectrum in your observations of a supermassive black hole, what might you conclude?
21. **How Do We Know?** How would you respond to someone who said about a scientific claim he or she didn't like, "Oh, that's only statistics"?

ACTIVE INQUIRY QUESTIONS

1. Why is it reasonable to assume that most large galaxies were once active galaxies?
2. Why do quasars, active galaxies, SS 433, and protostars have similar geometry?
3. Compare and contrast the origin of stellar-mass black holes versus supermassive black holes.
4. Astronomers normally refer to the unified *model* of AGN and not to the unified *hypothesis* or unified *theory*. In your opinion, which of these words seems best and why?
5. Could our Galaxy become an active galaxy in the future? How might that happen?
6. Why were quasars common when the Universe was young and not now?
7. A quasar has a redshift of 6. Is this galaxy near or far? When the light left the galaxy, was the galaxy young or old? Was the Universe young or at its current age? What was occurring on Earth when the light we see now left this quasar?

PROBLEMS

1. The total energy stored in a radio lobe is about 10^{53} J. How much mass would have to be converted to energy to produce that? Express your answer in units of solar mass. (*Hint:* Use $E = mc^2$.) (*Note:* One $M_\odot$ equals 2.0×10^{30} kg.)
2. If the jet in NGC 5128 is traveling at 5000 km/s and is 40 kpc long, how long will a gas blob take to travel from the nucleus of the galaxy to the end of the jet? (*Notes:* 1 pc equals 3.1×10^{13} km and 1 yr $= 3.2 \times 10^7$ s.)
3. Cygnus A is approximately 225 Mpc away, and its jet is about 50 arc seconds long. What is the length of the jet in units of parsecs? (*Hint:* Use the small-angle formula, Eq. 3-1.)
4. Find the linear diameter of a radio source that has an angular diameter of 0.0015 arc second and a distance of 3.25 Mpc. (*Hint:* Use the small-angle formula, Eq. 3-1.)
5. If the active nucleus of a galaxy contains a supermassive black hole with $10^6\ M_\odot$, what will the orbital period be for a blob of matter orbiting at a distance of 0.33 AU? (*Hint:* See the formula for circular velocity, Eq. 5-1a.) (*Notes*: 1 $M_\odot$ equals 2.0×10^{30} kg and 1 AU $= 1.5 \times 10^{11}$ m.)
6. If a quasar is 1000 times more luminous than an entire galaxy, what is the absolute magnitude of such a quasar? (*Hint:* See Table 2-1.) (*Note:* The absolute magnitude of a bright galaxy is about −21.)
7. If the quasar in Problem 6 were located at the center of the Milky Way Galaxy, what would its apparent magnitude be? (*Hints:* Use the magnitude–distance formula, Eq. 9-2b, and ignore dimming by dust clouds.)
8. Recalculate Problem 7 assuming the gas and dust located between Earth and the center of the Milky Way Galaxy dims the light by 30 magnitudes at visual wavelengths.
9. What is the change in the wavelength of the Balmer H-alpha line in the emission spectrum of 3C 273 due to the object's redshift? (*Hint:* Use the Doppler formula, Eq. 7-3.) (*Note:* The Balmer H-alpha line has a laboratory wavelength λ_0 of 486.1 nm.)
10. If the inner part of an accretion disk around a supermassive black hole is radiating at $T = 1$ million K, calculate the peak wavelength of the radiation. Express your answer in units of nm. Which part of the electromagnetic spectrum is this peak wavelength in? (*Hints:* Use Wien's law, Eq. 7-2, and refer to Figure 6-3.)
11. If the Hubble constant is 70 km/s/Mpc and a quasar has an apparent velocity of recession of 45,000 km/s, how far away is the quasar?

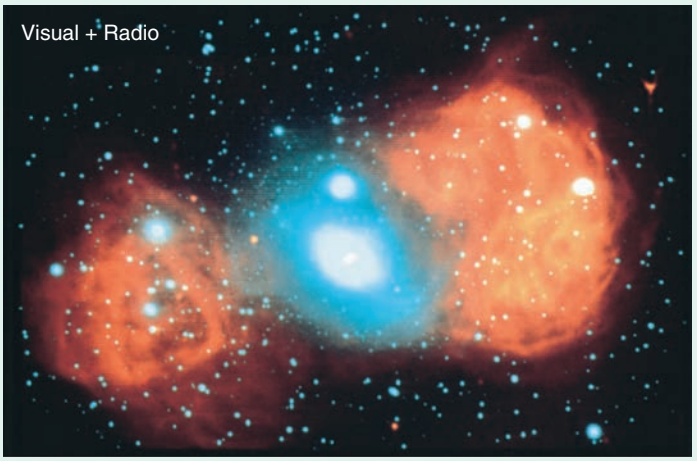
12. A quasar has a redshift of 6.0. What is the change in wavelength, $\Delta\lambda$, of the Balmer H-beta line? In which bands of the electromagnetic spectrum are the unshifted and shifted lines? Which quasar is farther from Earth, this one or 3C 273? (*Hints:* Use the Doppler formula, Eq. 7-3, and refer to Figure 6-3.) (*Note:* The Balmer H-beta line has a laboratory wavelength λ_0 of 486.1 nm.)
13. A quasar has a redshift of 6. What is its speed toward or away from Earth? Is the quasar moving toward or away from Earth? Express your answer in units of the speed of light *c*. (*Note:* The speed of light is 3.0×10^8 m/s.)

SENSE OF PROPORTION

1. Arrange the following in order of increasing size: a giant elliptical galaxy; our Milky Way Galaxy; a pair of radio lobes around an active galaxy; our Solar System

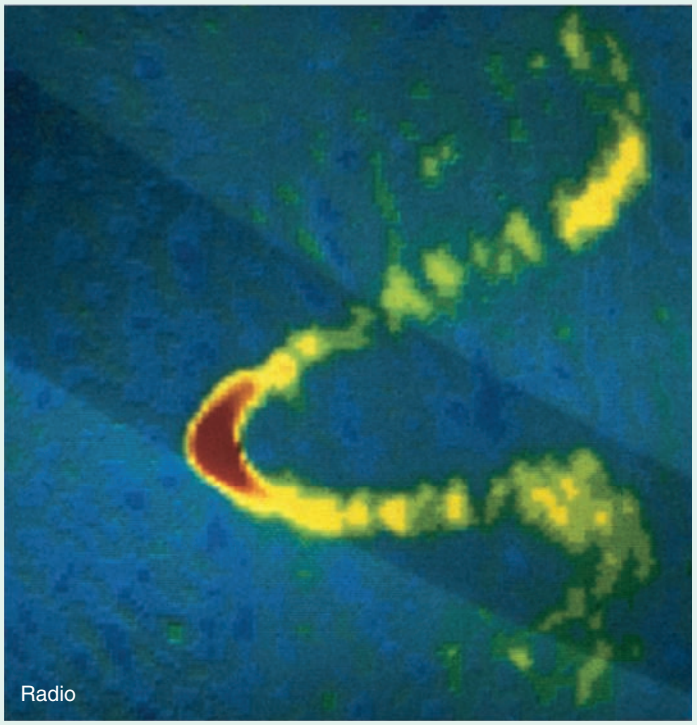
LEARNING TO LOOK

1. This image combines visual (*blue*) with radio (*orange-red*) to show the galaxy that radio astronomers call Fornax A. Explain the features of this image. Is it significant that the object is a distorted elliptical galaxy in a dense cluster of galaxies or is that just a coincidence?



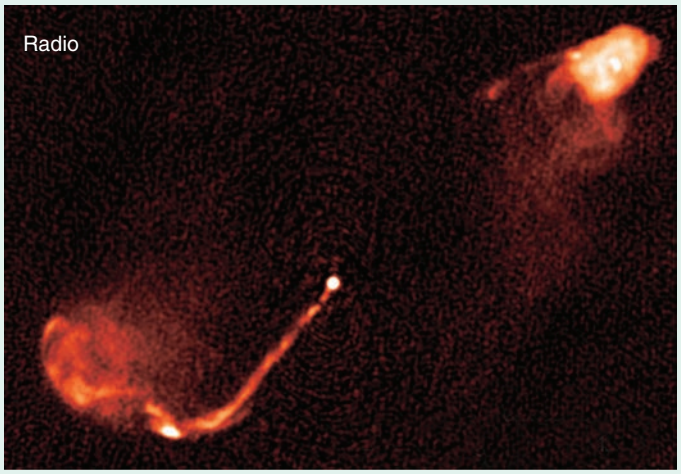
NRAO/AUI/NSF/J. Uson

2. Explain the features of the radio image of the galaxy IC 708, shown here.



NRAO/AUI/NSF

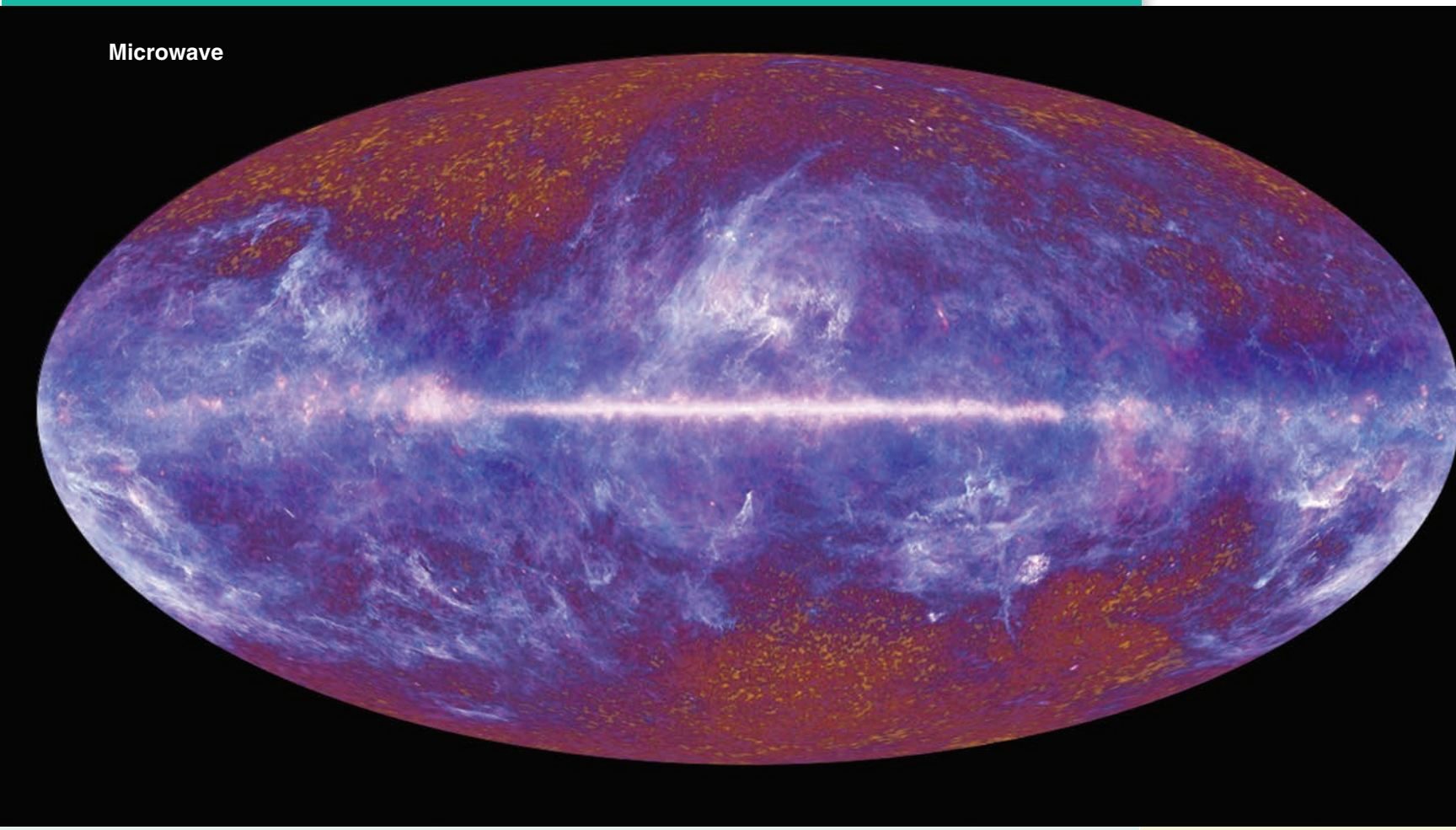
3. A radio image of quasar 3C 334 is shown here. Why do you see only one jet if two jets are known to be coming from the quasar? What does it mean that this looks so much like a radio image of a double-lobed radio galaxy?



NRAO/AUI/NSF

4. Look at Figure 17-6. Which wavelength is shown to be shifted—H-alpha, H-beta, H-gamma, or H-delta? Which spectral lines are shown to be shifted? Is the shift a redshift or a blueshift? How do you know?
5. Examine Figure 17-10. What was happening on Earth at the time the light from these quasars began its journey?

18 Modern Cosmology



ESA/*Planck*/HFI & LFI Consortia

▲ An all-sky map at microwave wavelengths made by the *Planck* space telescope. Purple colors represent foreground emission from gas and dust in the Milky Way Galaxy. The orange and red colors at top and bottom represent the cosmic microwave background, emitted about 400,000 years after the Big Bang.

You have been on an outward journey through the Universe since Chapter 1. Now, you have reached the limit of your travels in space and time and can study the Universe as a whole. The ideas in this chapter are among the biggest and most difficult in all of science. Can you imagine a limitless universe, or the first instant of time?

As you explore cosmology, you will find answers to three important questions:

- Does the Universe have an edge and a center?
- What is the evidence that the Universe began with a "Big Bang," an expansion from a hot, dense state?
- How has the Universe evolved, and what will be its fate?

Once you have finished this chapter, you will have a modern insight into the nature of the Universe, as well as where you are and what you are. After that, with an understanding of the big picture, you can consider how your local neighborhood—the Solar System—fits in.

> **The Universe, as has been observed before, is an unsettlingly big place, a fact which for the sake of a quiet life most people tend to ignore.**
>
> DOUGLAS ADAMS, *THE RESTAURANT AT THE END OF THE UNIVERSE*

LOOK AT YOUR THUMB. The matter in your body was present in the fiery beginning of the Universe. **Cosmology**, which is the study of the Universe as a whole, can tell you where your atoms came from, and it can tell you where your atoms are going.

Cosmology is a mind-bendingly weird subject, and you can enjoy it for its strange ideas. It is fun to think about the origin of vast walls of galaxy clusters, about space stretching like a rubber sheet, and about invisible energy pushing the Universe to expand faster and faster. Notice that this is better than just speculation; all of these concepts are supported by evidence. Cosmology, however strange it may seem, is a careful and logical attempt to understand the structure and evolution of the entire Universe.

This chapter will help you climb the cosmology pyramid one step at a time (**Figure 18-1**). You already have some ideas about what the Universe is like. Start with those ideas, test them against observations, and compare them with scientific hypotheses. Step by step you can build a modern understanding of cosmology. Each step in the pyramid is small, but it leads to some astonishing insights into how the Universe works and how you came to be a part of it.

▲ **Figure 18-1** Climbing the cosmology pyramid step by step isn't difficult, and it leads to some fascinating ideas about the origin and evolution of the Universe.

18-1 Introduction to the Universe

Most people have an impression of the Universe as a vast ocean of space filled with stars and galaxies (**Figure 18-2**), but as you begin exploring the Universe, you need to become aware of your expectations so they do not mislead you. The first step is to deal with an expectation so obvious that most people don't think about it, for the sake of a quiet life.

18-1a The Edge–Center Problem

In your daily life, you are accustomed to boundaries. Rooms have walls, athletic fields have boundary lines, countries have borders, and oceans have shores. It is natural to think of the Universe also as having an edge, but that idea can't be right.

If the Universe had an edge, imagine going to that edge. What would you find there: A wall of some type? A great empty space? Nothing? Even a child can ask: If there is an edge to the Universe, what's beyond it? A true edge would have to be more than just an end of the distribution of matter. It would have to be an end of space itself. But, then, what would happen if you tried to reach past, or move past, that edge?

An edge to the Universe violates common sense. Perhaps even more important, the centers of things—pizzas, football fields, oceans, and galaxies—are found by referring to their edges. If the Universe has no edge, then it cannot have a center. It is a **Common Misconception** to imagine that the Universe has

NASA/ESA/STScI/AURA/R. Williams, HDF-South Team

▲ **Figure 18-2** The entire sky is filled with galaxies. Some lie in clusters of thousands, and others are isolated in nearly empty voids between the clusters. In this image of a typical spot on the sky, most of the bright objects are foreground stars; their "spikes" are caused by diffraction in the telescope. All other objects are galaxies ranging from the relatively close face-on spiral at upper right to the most distant galaxies visible only in the infrared, shown as red in this composite image.

a center, but now you understand that is impossible. As you study cosmology, you should take care to avoid thinking that the Universe has a center or an edge.

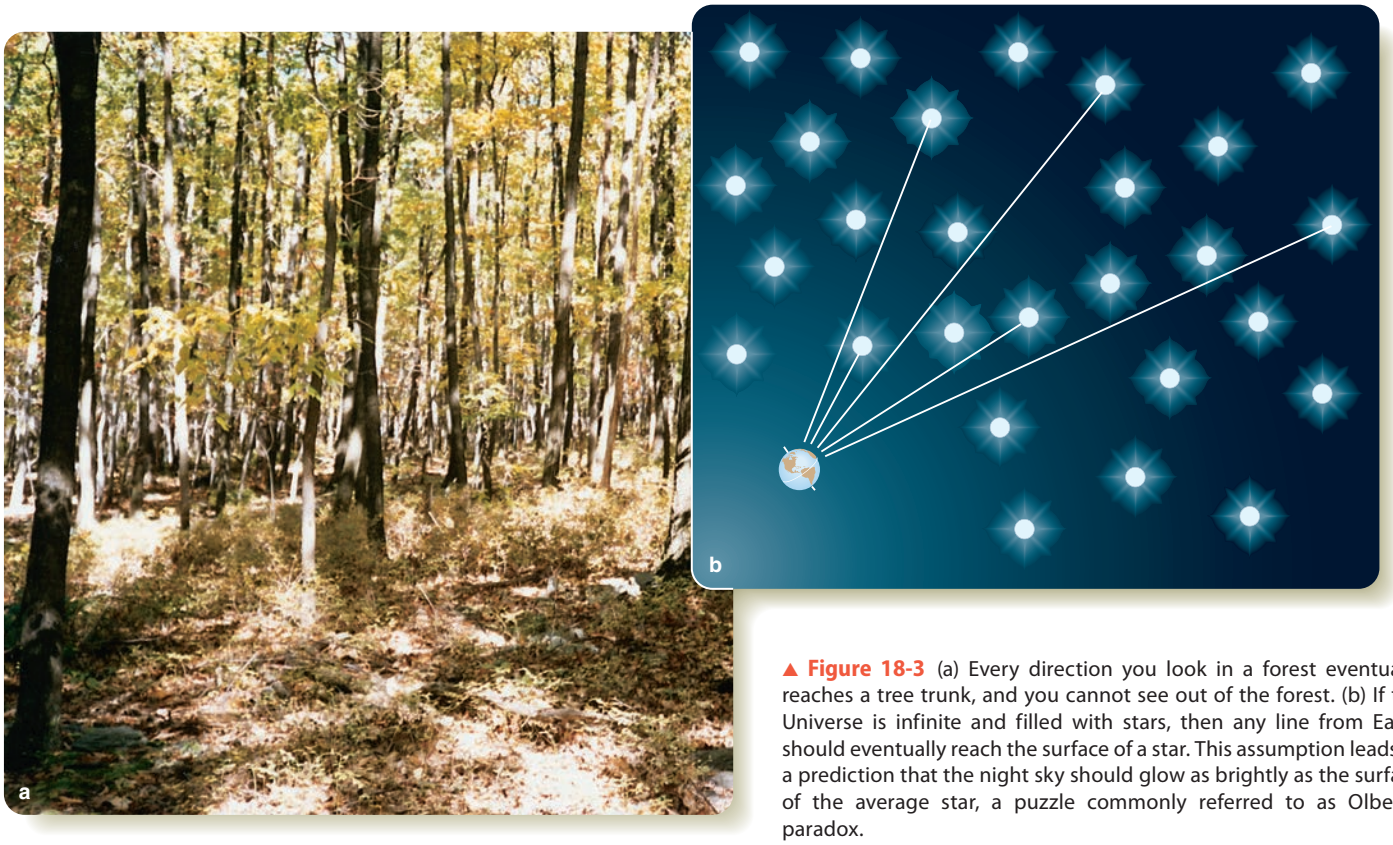
18-1b The Idea of a Beginning

Of course you have noticed that the night sky is dark. That is an important observation because, you may be surprised to learn, seemingly reasonable assumptions about the Universe can lead to the conclusion that the night sky actually should glow blindingly bright. This conflict between observation and theory is called **Olbers's paradox** after Heinrich Olbers, a physician and astronomer who publicized the problem in 1826. (*Problem* or *question* might be more accurate words than *paradox*.) The question of why the night sky is dark was first discussed by Thomas Digges in 1576, but Olbers gets the credit because of incomplete scholarship on the part of modern scientists who were not aware of previous discussions.

The point made by Olbers seems simple. Suppose you assume, as did most scientists in Olbers's time, that the Universe is infinite in size, infinite in age, *static* (a fancy word for unchanging overall), and filled with stars. If you look in any direction, your line of sight must eventually run into the surface of a star. (The clumping of stars into galaxies and galaxies into clusters can be shown mathematically to make no difference.)

Look at **Figure 18-3**, which uses the analogy of lines of sight in a forest. (The use of analogies in science is discussed in **How Do We Know? 18-1**.) When you are deep in a forest, every line of sight ends at a tree trunk, and you cannot see out of the forest. By analogy to the view from inside a forest, every line of sight from Earth into space should eventually end at the surface of a star. Of course, the more distant stars would be fainter than nearby stars because of the inverse square law. However, the farther you look into space, the larger the volume you are viewing and the more stars are included; the two effects cancel out. The result would be that the entire sky should be as bright as the surface of an average star—like suns crowded "shoulder to shoulder," covering the sky from horizon to horizon. It should not get dark at night.

Scientists who study the Universe as a whole, called **cosmologists**, now are sure they understand why the sky is dark. Olbers's paradox is based on assumptions that we now know are incorrect. The Universe may be infinite in size, but it is neither infinitely old nor static. The essence of modern cosmologists' answer to Olbers's paradox was suggested first by Edgar Allan Poe in 1848. Poe proposed that the night sky is dark because the Universe came into existence at some time in the past and therefore has a finite age. As a consequence, if you look far enough away, the look-back time (look back to Section 16-2a) becomes almost equal to the age of the Universe. You can see back to a time before the first stars began to shine. There may be stars more distant than that, but their light has



Janet Seeds

▲ **Figure 18-3** (a) Every direction you look in a forest eventually reaches a tree trunk, and you cannot see out of the forest. (b) If the Universe is infinite and filled with stars, then any line from Earth should eventually reach the surface of a star. This assumption leads to a prediction that the night sky should glow as brightly as the surface of the average star, a puzzle commonly referred to as Olbers's paradox.

Reasoning by Analogy

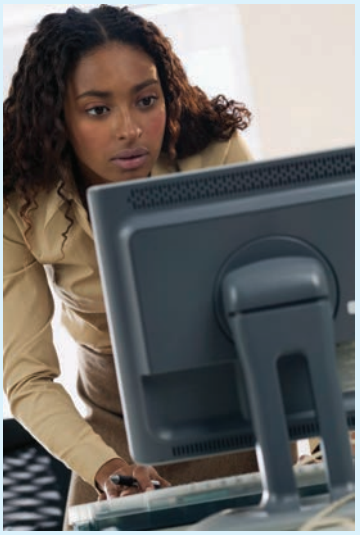
How do scientists use analogies?

"The economy is overheating, and it may seize up," an economist might say. Economists like to talk in analogies because economics is often abstract, and one of the best ways to think about abstract problems is to find a more approachable analogy. Rather than discussing details of the national economy, you might be able to make conclusions about how the economy works by thinking about how a gasoline engine works.

Much of astronomy is abstract, and cosmology is the most abstract subject in astronomy. Furthermore, cosmology is highly mathematical, and unless you are prepared to learn some difficult mathematics, you instead have to use analogies, such as lines of sight in a forest.

Reasoning by analogy is a powerful technique. An analogy can reveal unexpected insights and lead you to further discoveries. Carrying an analogy too far, however, can be misleading. You might compare the human brain to a computer, and that would help you understand how data flow in and are processed and how new data flow out, but the analogy is flawed. For example, although data in computers are stored in specific locations, memories are stored in the brain in a distributed form. No single brain cell holds a specific memory. So if you carry the analogy too far, it can mislead you. Whenever you reason using analogies, you should be alert for their limitations.

As you study any science, be alert for analogies. They are tremendously helpful, but you have to be careful not to carry them too far.



Blend Images/Superstock

The analogy between a human brain and a computer is of only limited use.

not yet reached Earth. By revising their original assumptions about the Universe, cosmologists can now answer Olbers's paradox and understand why the night sky is dark: because the Universe had a beginning.

The answer to Olbers's question is a powerful idea because it clearly illustrates the difference between the Universe and the **observable universe**. The Universe is everything that exists and could be infinite. The observable universe, in contrast, is the part (possibly a very small part) that you can see from Earth using the most powerful telescopes. You will learn later in this chapter about evidence that the Universe is approximately 14 billion years old. That means you can't observe objects farther away than a look-back time of about 14 billion years. Be careful not to confuse the observable universe, which is huge but finite, with the Universe as a whole, which might be infinite.

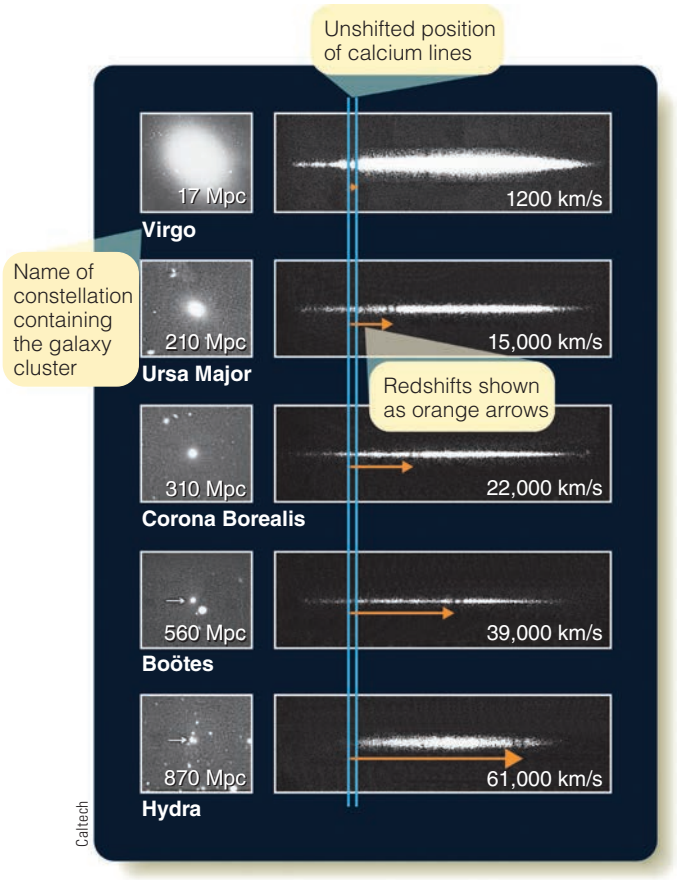
18-1c Cosmic Expansion

There is a **Common Misconception** that the Universe is unchanging overall. Having tackled the notion that the Universe had a beginning, you are ready to understand evidence that the Universe actually continues to change and evolve.

In 1929, Edwin Hubble published his discovery that the sizes of galaxy redshifts are proportional to galaxy distances. Nearby galaxies have small redshifts, and more distant galaxies have large redshifts. You learned this as the Hubble law when you used it to estimate the distances to galaxies (Eq. 16-1). Those galaxy redshifts imply that galaxies are receding from each other. That idea has become known as the **expanding universe**.

Figure 18-4 shows spectra of galaxies in galaxy clusters at various distances. The Virgo cluster is relatively nearby, and its redshift is small. The Hydra cluster is very distant, and its redshift is so large that the two dark spectral lines formed by ionized calcium absorption are shifted from near-ultraviolet wavelengths well into the visible part of the spectrum.

The expansion of the Universe does not imply that Earth has a special location. To see why, look at **Figure 18-5**, which shows an analogy to baking raisin bread. As the dough rises, it pushes the raisins away from each other at speeds that are proportional to distance. Two raisins that were originally close to each other are pushed apart slowly, but two raisins that were far apart, having more dough between them, are pushed apart faster. If bacterial astronomers lived on a raisin in the raisin bread, they could observe the redshifts of the other raisins and derive a bacterial Hubble law. They would conclude that their universe was expanding uniformly. It does not matter which raisin the bacterial astronomers lived on, they would get the same Hubble law—no raisin has a special viewpoint. Similarly,



▲ **Figure 18-4** These galaxy spectra extend from near-ultraviolet wavelengths at left to the blue part of the visible spectrum at right. The two dark absorption lines of once-ionized calcium are prominent in the ultraviolet. The redshifts in galaxy spectra are expressed here as apparent velocities of recession. Note that the apparent velocity of recession is proportional to distance, which is known as the Hubble law.

astronomers in any galaxy will see the same law of expansion—no galaxy has a special viewpoint.

When you look at Figure 18-5, you see the edge of the loaf of raisin bread, and you can identify a center to the loaf. The raisin bread analogy to the Universe doesn't work when you consider the crust (the edge) of the bread. Remember that the universe cannot have an edge or a center, so there is no center to the expansion. The raisin bread analogy is useful but also imperfect.

18-2 The Big Bang Theory

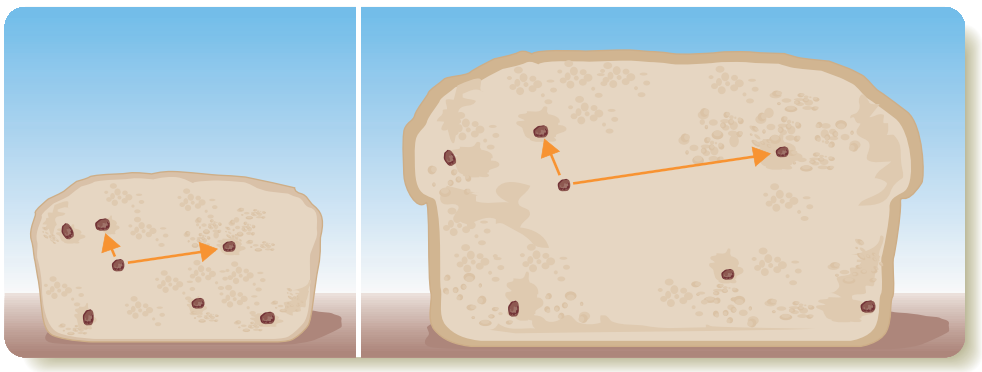
Now you are ready to take a historic step up the cosmology pyramid. The expansion of the Universe led cosmologists to conclude that the Universe must have begun with an event of unimaginable intensity. Popular television shows notwithstanding, you will learn that the Big Bang can be called a "theory" rather than a hypothesis because the supporting evidence is so solid and comprehensive.

18-2a Necessity of the Big Bang

Imagine that you have a video of the expanding Universe and run it backward. You would see the galaxies getting closer to each other. There is no center to the expansion of the Universe, so you would not see galaxies approaching a single spot. Rather, you would see galaxies "holding still" while the spaces between the galaxies shrank and the distances between all galaxies decreased. Eventually, as your video ran farther back, galaxies would begin to merge. If you ran the video far enough back, you would see the matter and energy of the Universe compressed into a high-density, high-temperature state. This thought experiment is the reason why cosmologists infer that the Universe must have expanded from a moment of extreme conditions that is called the **Big Bang**.

How long ago did the Universe begin? You can estimate the age of the Universe with a simple calculation. If you need to drive to a city 100 miles away, and you can travel 50 miles per hour, you divide distance by rate of travel and learn the travel time—in this example, 2 hours. In a similar fashion, to find the age of the Universe, you can divide the distance between two galaxies by the speed with which they are moving apart and find out how much time they have taken to reach their present separation.

A more general way to make the calculation is to use the Hubble constant, which summarizes the velocities and separations of all galaxies. The Hubble constant, H_0, has units of km/s per Mpc, which is a velocity divided by a distance. If you calculate the reciprocal of the Hubble constant, $1/H_0$, you have a distance divided by velocity, which is a time. To actually perform the division and get a quantity in time units, you need



◀ **Figure 18-5** An illustration of the raisin bread analogy for the expansion of the Universe. As the dough rises, raisins are pushed apart with velocities proportional to distance. A colony of bacteria living on any raisin will find that the redshifts of the other raisins are proportional to their distances.

to convert megaparsecs into kilometers, and then the distance units will cancel properly and leave you with an age in seconds. To convert seconds into years, you divide by the number of seconds in a year. If you make these simple changes in units, the age of the Universe τ_H in years is approximately 10^{12} divided by H, if H is expressed in the units astronomers conventionally use, km/s/Mpc:

KEY EQUATION 18-1 The Age of the Universe

$$\tau_H = \frac{10^{12}}{H_0} \text{ years}$$

This estimate of the age of the Universe is known as the **Hubble time**. ***Example:*** If H_0 is 70 km/s/Mpc, what is the estimated age of the Universe? ***Solution:*** The estimated age of the Universe is:

$$\frac{10^{12}}{70} \text{ years}$$

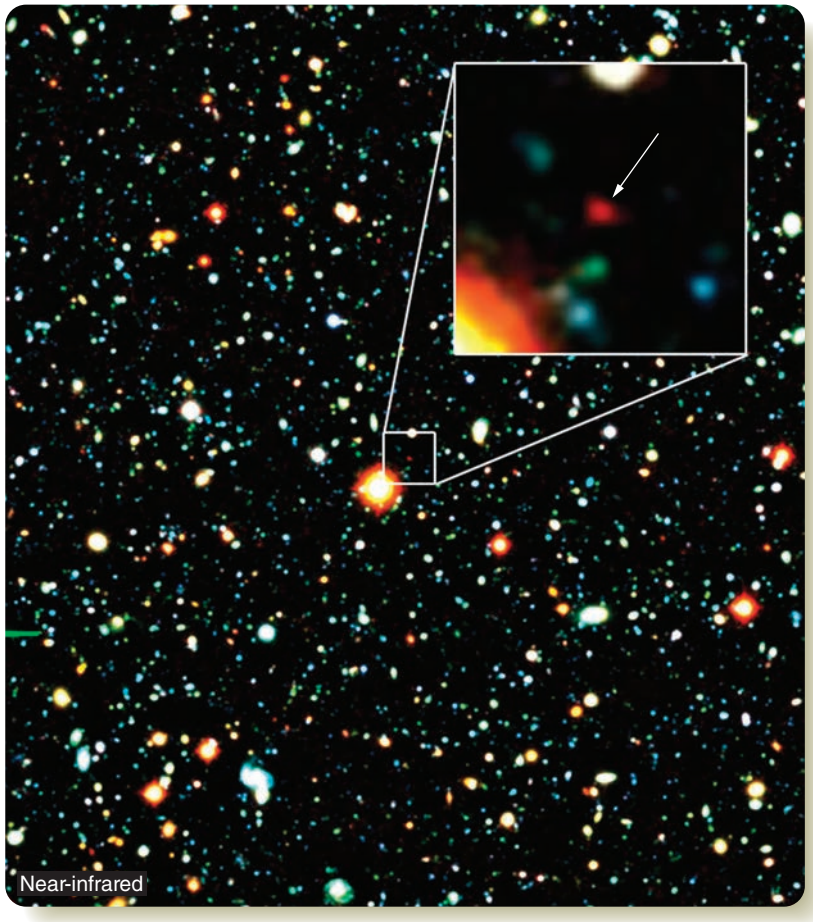
or about 14 billion years. This estimate of the age of the Universe will be fine-tuned later in this chapter, but for the moment you can conclude that basic observations of the recession of the galaxies require that the expansion of the Universe began about 14 billion years ago.

The phrase *big bang* was invented by early critics of the original hypothesis, and the label gives a misimpression. Do not think of an edge or a center when you think of the Big Bang. It is a **Common Misconception** that the Big Bang was an explosion and that the galaxies are flying away from the location of that explosion. Instead, you should try to keep firmly in mind the correct picture that the Big Bang did not occur at a single place but filled the entire volume of the Universe. A more accurate term than *big bang* might be *big stretch.* The matter of which you are made was part of that Big Bang, so you are inside the remains of that event, and the Universe continues to expand around you. You cannot point to any particular place and say, "The Big Bang occurred over there." This is a brain-straining (stretching?) idea, but you will find more information later in this chapter to help you understand it when you study the nature of space and time.

You might imagine the Big Bang as an event that happened long ago and can no longer be observed, like the Gettysburg Address. Amazingly, the effect of look-back time makes it possible to observe the early Universe now, directly. The look-back time to nearby galaxies is "only" a few million years; the look-back time to more distant objects is a large fraction of the age of the Universe (**Figure 18-6**).

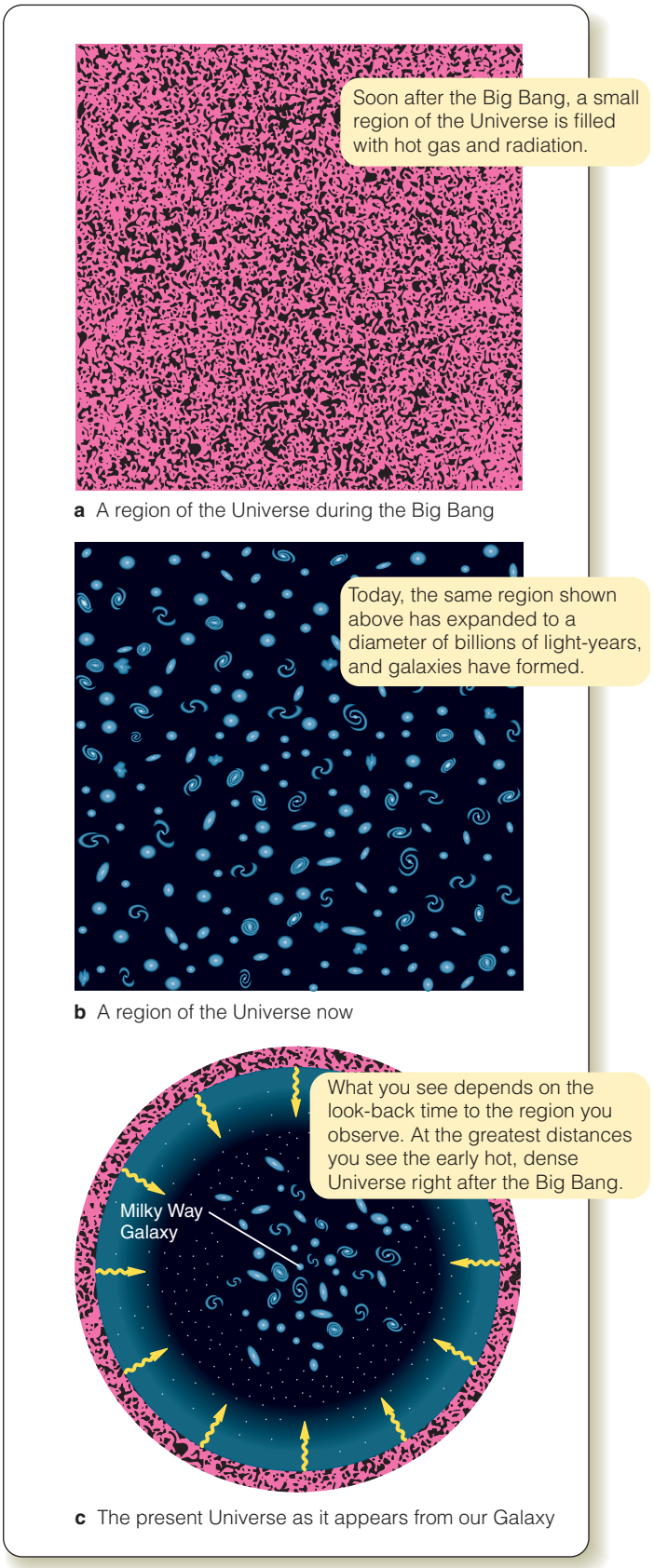
Suppose you look between the distant galaxies, seeing even farther away and farther back in time. You should be able to detect the hot gas that filled the Universe long ago, right after the Big Bang, before the first stars and galaxies formed. The Big Bang occurred everywhere, so that, in whatever direction you look, at great distance you can see back to the age when the Universe was filled with hot gas (**Figure 18-7**).

Radiation that comes from such a great distance must have a large redshift. The most distant objects known are some galaxies and quasars with redshifts less than 10. Radiation emitted by the hot gas right after the Big Bang, long before galaxies and quasars formed, must have a much larger redshift that that, so you might guess that it might be detectable at long wavelengths, using infrared and radio telescopes. In fact, unlike the Gettysburg Address, the Big Bang can still be observed. That amazing discovery is the subject of the next section.



NAOJ/Subaru Telescope. All rights reserved.

▶ **Figure 18-6** This faint galaxy is one of the most distant ever found. It has a redshift of 6.96, implying that the observed light left this source 12.9 billion years ago. It appears as it was only 850 million years after the Big Bang when its light began its journey toward Earth.


Soon after the Big Bang, a small region of the Universe is filled with hot gas and radiation.
Today, the same region shown above has expanded to a diameter of billions of light-years, and galaxies have formed.
What you see depends on the look-back time to the region you observe. At the greatest distances you see the early hot, dense Universe right after the Big Bang.
Milky Way Galaxy

a A region of the Universe during the Big Bang

b A region of the Universe now

c The present Universe as it appears from our Galaxy

◀ **Figure 18-7** This diagram shows schematically the expansion of a small part of the Universe. Although the Universe is now filled with galaxies, the look-back time distorts what you see. Nearby you see galaxies, but at greater distances the look-back time reveals the Universe at earlier stages, before galaxies formed. At very great distances, the hot gas that filled the Universe soon after the Big Bang is detectable as a source of infrared, microwave, and radio energy arriving from all directions.

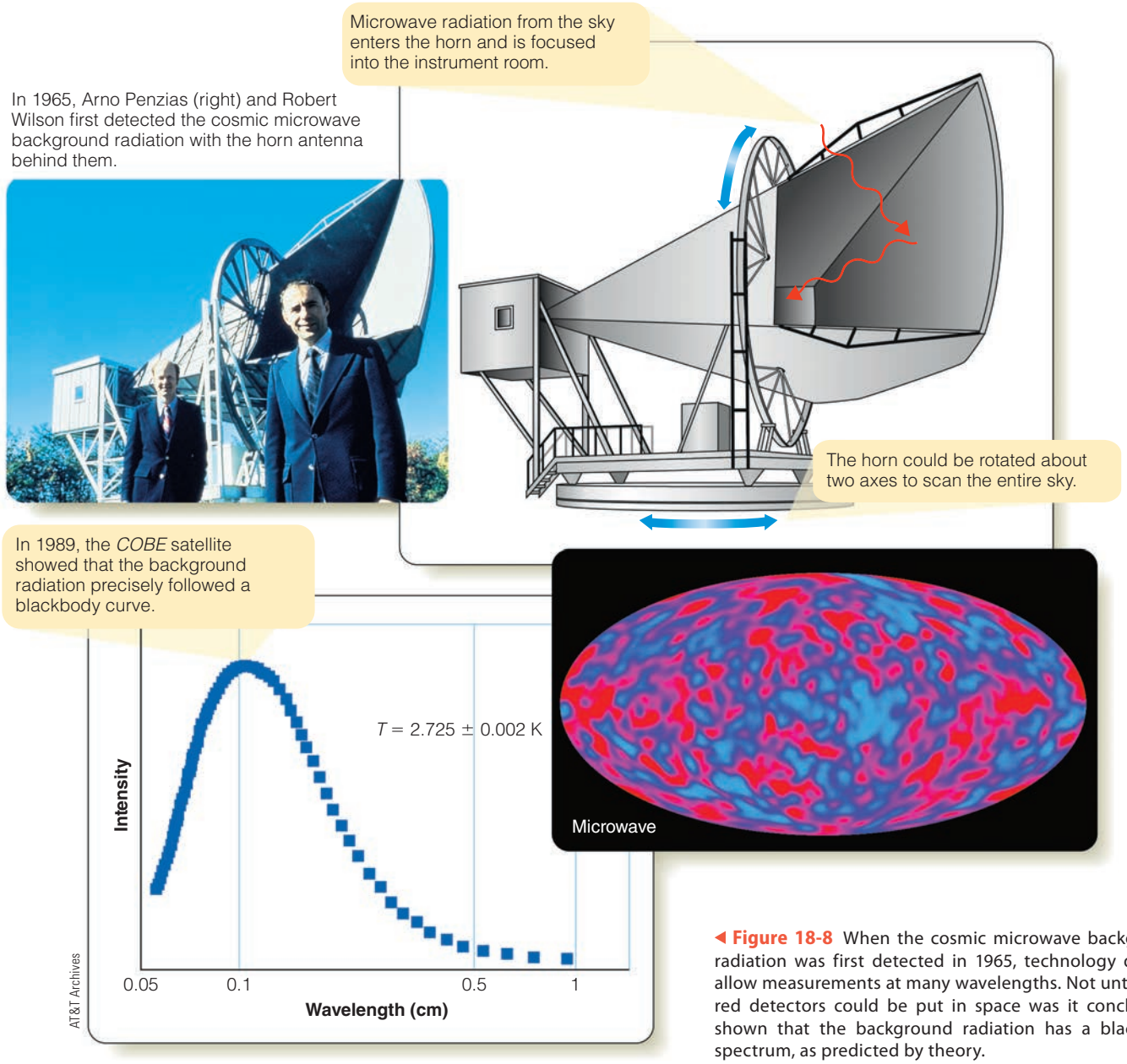
18-2b The Cosmic Background Radiation

The story of the discovery of radiation from the time of the Big Bang begins in the mid-1960s when two Bell Laboratories physicists, Arno Penzias and Robert Wilson, were measuring the brightness of the sky at radio wavelengths (**Figure 18-8**). Their measurements showed a strange extra signal in the system, which they first attributed to the infrared glow from pigeon droppings inside the antenna. Perhaps they would have enjoyed scraping out the antenna more if they had known they would win the 1978 Nobel Prize in physics for the discovery they were about to make.

When the antenna was cleaned, they again measured the radio brightness of the sky and found the low-level noise was still there. The pigeons were innocent, but what was causing the extra signal?

The explanation for the noise goes back decades earlier. In 1939, astronomers noticed that spectra of some molecules in the interstellar medium showed they were bathed in radiation from a source with a temperature of 2 to 3 K. In 1948, physicist George Gamow predicted that the gases in the Universe right after the Big Bang would have been hot and should have emitted strong blackbody radiation (Section 7-2). A year later, physicists Ralph Alpher and Robert Herman pointed out that the large redshift of the Big Bang material relative to Earth would lengthen the wavelengths of that radiation into the microwave part of the spectrum, with an observable blackbody temperature they estimated as 5 K. In the mid-1960s, Robert Dicke at Princeton concluded that this radiation should be just strong enough to detect with newly developed techniques, so he and his team began building a receiver. When Penzias and Wilson heard of Dicke's work, they recognized the mysterious extra signal they had detected as radiation from the Big Bang, the **cosmic microwave background radiation**.

The detection of the background radiation was tremendously exciting, but cosmologists wanted confirmation. Theory predicted that the radiation should have a spectrum like blackbody radiation coming from a very cool source, but the critical observations could not be made from the ground because Earth's atmosphere is opaque at the predicted blackbody peak wavelength. It was not until 1989 that satellite measurements confirmed the background radiation has exactly a blackbody spectral distribution, with an apparent temperature of 2.725 ± 0.002 K—close to the original prediction.


Microwave radiation from the sky enters the horn and is focused into the instrument room.
In 1965, Arno Penzias (right) and Robert Wilson first detected the cosmic microwave background radiation with the horn antenna behind them.
The horn could be rotated about two axes to scan the entire sky.
In 1989, the COBE satellite showed that the background radiation precisely followed a blackbody curve.
T = 2.725 ± 0.002 K
Intensity
0.05
0.1
0.5
1
Wavelength (cm)
Microwave
AT&T Archives

◀ **Figure 18-8** When the cosmic microwave background radiation was first detected in 1965, technology did not allow measurements at many wavelengths. Not until infrared detectors could be put in space was it conclusively shown that the background radiation has a blackbody spectrum, as predicted by theory.

It may seem strange that the gas of the Big Bang seems to have a temperature of just 2.7 K above absolute zero, but recall the tremendous redshift. The gas clouds that originally emitted those photons would have had a temperature calculated to be about 3000 K, so emitting blackbody radiation with a λ_{max} of about 1000 nm (Wien's law, Eq. 7-2). (Although that wavelength is in the near-infrared, the gas would also have emitted enough visible light to be seen glowing orange-red if there had been a human eye present at the time.) Observers on Earth now receive cosmic background radiation with λ_{max} of about 1 mm, or 1 million nm, in the microwave portion of the electromagnetic spectrum. This represents a redshift of about 1000—that is, the wavelengths of the received photons are 1000 times longer than when they were emitted. That is why that hot gas seems to be about 2.7 K, a factor of about 1000 cooler than it actually was.

The first few steps up the cosmology pyramid have not been very difficult. Simple observations of the darkness of the night sky and the redshifts of the galaxies tell you that the Universe must have had a beginning, and the cosmic microwave background radiation is clear evidence that the early Universe was hot and dense, the Big Bang. Theorists can combine these observations with modern physics to add some details to the story of how the Big Bang occurred.

18-2c Photon and Particle Soup

Cosmologists cannot begin their history of the Big Bang at a time of exactly zero because no one understands the physics of matter and energy under such extreme conditions, but they can understand what was happening surprisingly close to time zero by making hypotheses and checking them with observations. Cosmologists calculate that, if you could have visited the Universe when it was only one-millionth of a second old, you would have found it filled with high-energy photons having a temperature of 20 trillion (2×10^{13}) K. (When cosmologists say the photons have a given temperature, they mean the photons have the same spectrum as blackbody radiation emitted by an object of that temperature.) Consequently, the photons in the early Universe were gamma-rays, with very short wavelengths and very high energies. Moreover, the same calculations indicate that when the Universe was one-millionth of a second old, the density of the radiation was 5×10^{20} kg/m^3, more than a thousand times the density of an atomic nucleus. (When cosmologists say that the radiation had a certain density, they refer to Einstein's equation $E = mc^2$. Using that equation, you can express a given amount of radiation energy per volume as if it were matter of a certain density.)

If photons have enough energy, two photons can convert into a pair of particles—a particle of ordinary matter and a particle of **antimatter**. On the other hand, when an antimatter particle meets its matching particle of ordinary matter—when an antiproton meets a proton, for example—the two particles annihilate each other and convert their mass into energy in the form of two gamma-rays. Early enough in the history of the Universe, the most abundant photons were gamma-rays with enough energy to produce proton–antiproton pairs or neutron–antineutron pairs. When these particles collided with their antiparticles, they converted their mass back into photons. Thus, the early Universe was filled with a dynamic soup of energy flickering from photons into particles and back again.

While all this went on, the expansion of the Universe caused the temperature of the radiation to drop, reducing the energy of the photons. Cosmologists can calculate, based on the known properties of subatomic particles and also the characteristics of the Universe as a whole, that when the Universe was ten-millionths of a second old, its temperature had fallen to about 2 trillion (2×10^{12}) K. By that time, the average energy of the radiation photons had fallen well below the energy equivalent to the mass of proton–antiproton or neutron–antineutron pairs, so the gamma-rays could no longer produce such heavy particles. The particles in existence at that time combined with their antiparticles and quickly converted their mass into photons that, as the Universe expanded and increased their wavelengths, could not later convert back into massive particles.

It would seem that all of the protons and neutrons should have been annihilated with their antiparticles, but, for reasons that are poorly understood, a small excess of ordinary particles existed. For every billion protons annihilated by antiprotons, one proton survived with no antiparticle to destroy it. Consequently, you live in a world of matter, and antimatter is very rare.

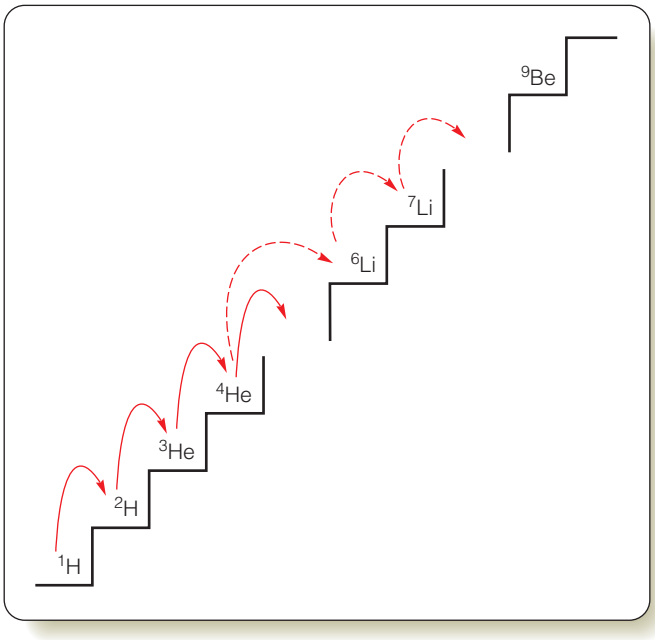
Although the gamma-ray photons did not have enough energy to produce any more protons and neutrons after the Universe was older than about one-millionth of a second, they could still produce electrons and antielectrons (called positrons; Section 8-3b) because those particles are about 1800 times less massive—require 1800 times less energy to create—than protons and neutrons. Electron–positron production continued until the Universe had expanded and cooled to the point at which there were no remaining photons with enough energy to create electron–positron pairs. Then, similar to the end of the proton–antiproton era, most of the electrons and positrons combined to form photons, and only one in a billion electrons survived. Cosmologists calculate that the electron–positron era ended when the Universe was about 1 minute old; this means that almost all the protons, neutrons, and electrons of which the Universe is now made were produced during the first minute of its history.

18-2d A Few Minutes of Nucleosynthesis

As the universal soup of hot gas and radiation expanded, it continued to cool. Photons with high enough energy can break up an atomic nucleus, so the formation of stable nuclei could not occur until the Universe had cooled below a certain temperature. By the time the Universe was about 2 minutes old, protons and neutrons could link to form deuterium (the nucleus of a heavy hydrogen atom) without being immediately broken apart. By the end of the third minute, further reactions began converting deuterium into helium.

Almost no atoms heavier than helium could be built in the Big Bang because there are no stable nuclei with atomic weights of 5 or 8 (in units of the hydrogen atom). Nuclei of atomic weights 5 and 8 are radioactive and decay almost instantly back into smaller nuclei. Cosmic element building during the Big Bang had to proceed step by step, like someone hopping up a flight of stairs (**Figure 18-9**). The lack of stable nuclei at atomic weights of 5 and 8 meant there were missing steps in the stairway, and the step-by-step reactions had great difficulty jumping over these gaps during the few minutes of the Big Bang. As a result, cosmologists can calculate that only a tiny amount of lithium (atomic weight 7) would have been produced during the Big Bang, and no heavier elements. Formation of elements with atomic weights greater than lithium had to wait for relatively slow-cooking nucleosynthesis processes in stars (Section 12-2c), beginning many millions of years after the Big Bang.

By the time the Universe was 3 minutes old, it had become cool enough that almost all nuclear reactions were



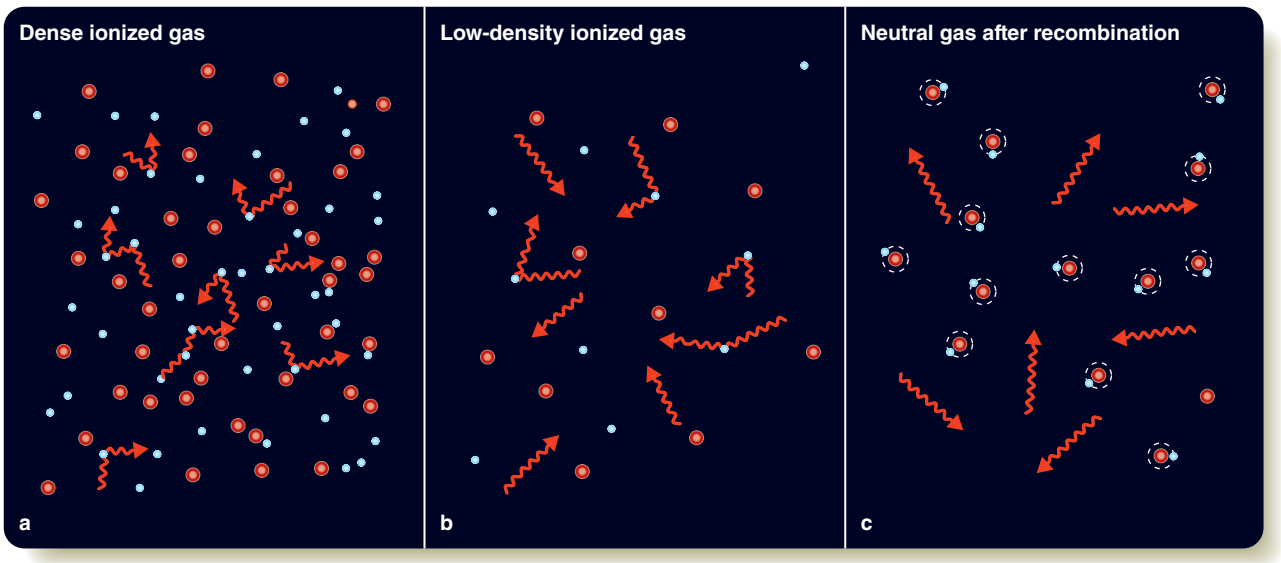
▲ **Figure 18-9** Cosmic element building: During the first few minutes of the Big Bang, temperatures and densities were high, and nuclear reactions built heavier elements. Because there are no stable nuclei with atomic weights of 5 or 8, the process built very few atoms heavier than helium.

slowing down. By the time it was 30 minutes old, the nuclear reactions had ended completely, and about 25 percent of the mass of the Universe was in the form of helium nuclei. The rest was in the form of protons—hydrogen nuclei. The fact that the composition observed for the oldest stars is exactly the composition that nuclear physics predicts would be produced by the Big Bang is the third major independent piece of evidence, in addition to the Hubble law and the cosmic background radiation, supporting the Big Bang theory.

18-2e Radiation and Matter: Recombination and Re-ionization

After the era of nucleosynthesis, as the young Universe continued to expand and cool, it went through a further sequence of four important changes. Originally, the Universe was so hot that the gas was totally ionized, and the electrons were not attached to nuclei. The free electrons interacted with photons so easily that a photon could not travel very far before it encountered an electron and was deflected (**Figure 18-10a**). Because photons interacted continuously with matter, the radiation and matter cooled together at a rate set by the expansion rate of the Universe.

Cosmologists calculate that the next big change occurred when the Universe reached an age of about 50,000 years. Then, the density of energy in the form of photons became less than the density of the gas. Before that time, the Universe was dominated by radiation; ordinary matter could not clump together because the intense sea of photons smoothed the gas out. As you will learn in the next section, dark matter, which does not interact with photons, was able to start clumping much earlier. After the 50,000-year mark, when the density of the Universe's radiation permanently became less than that of ordinary matter (made of protons, neutrons, and electrons), ordinary matter could begin to draw together under the influence of the



▲ **Figure 18-10** (a) Photons (*red waves*) scatter from electrons (*blue dots*) easily but hardly at all from the more massive but much smaller protons (*red dots; not to scale*). When the Universe was very dense and ionized, photons could not travel very far before they scattered off an electron. This made the gas opaque. (b) As the Universe expanded, the electrons were spread farther apart, and a photon could travel farther before encountering one; this made the gas more transparent. (c) After recombination, most electrons coupled with nuclei to form neutral atoms, and the gas became essentially completely transparent.

gravitational attraction of dark matter to form the clouds that eventually became galaxies.

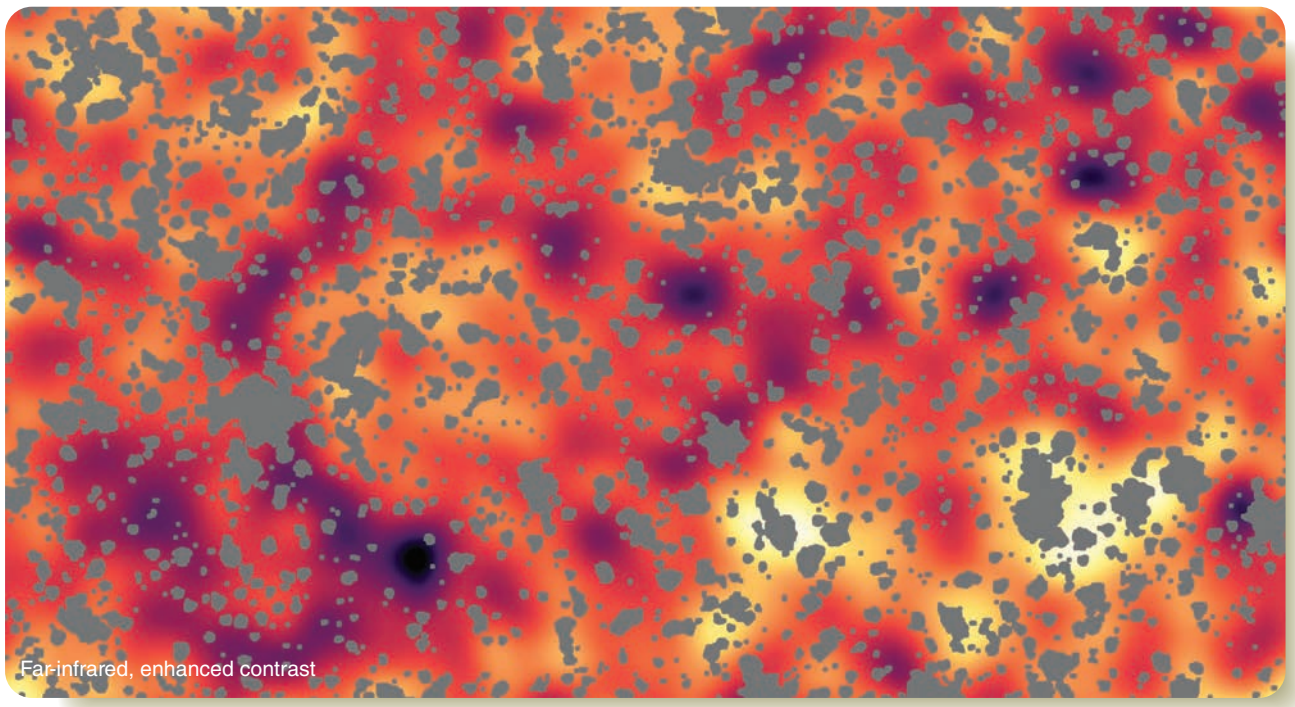
The expansion of the Universe spread the particles of the still-ionized gas farther and farther apart. When the Universe reached an age of about 400,000 years, the second important change began. By then, the free electrons were spread so far apart that photons could travel for thousands of parsecs before being deflected (**Figure 18-10b**). In other words, the Universe started to become transparent. At approximately the same time, the third important change happened. As the falling temperature of the Universe reached 3000 K, protons were able to capture and hold free electrons to form neutral hydrogen atoms, a process called **recombination**. (That term is a little misleading because the particles had never been able to hold together stably before; *combination* would be more accurate.) As the free electrons were gobbled up, the gas became close to completely transparent, and the photons could travel through the gas without being deflected (**Figure 18-10c**).

After recombination, although the gas continued to cool, the photons no longer interacted with the gas, and consequently the photons retained the blackbody temperature that the gas had at recombination. Those photons, which started their journey with a blackbody temperature of 3000 K, are observed now as the cosmic microwave background radiation. As you learned in the previous section, the expansion of the Universe has stretched the wavelengths of the background radiation so that it appears today to have a temperature of about 2.7 K.

After recombination, the gas that originated in the Big Bang was neutral, warm, and transparent. As the Universe continued to expand and cool, the glow from the warm gas faded into infrared wavelengths. The Universe entered what cosmologists call the **dark age**, a period lasting for an estimated 400 million years until the formation of the first stars. During that era, the Universe expanded in darkness.

The fourth remarkable change happened as the first stars began to form, ending the dark age. The gas from which the first stars formed contained almost no metals and was therefore highly transparent. Mathematical models show that stars forming from this metal-poor gas would have been very massive, very luminous, and very short lived. That first violent burst of massive star formation produced enough ultraviolet light to begin ionizing the gas, and astronomers now, looking back beyond and before the most distant visible quasars and galaxies, can detect evidence of that **re-ionization** era in the Universe (**Figure 18-11**). Re-ionization marked the end of the dark age and the beginning of the age of stars and galaxies in which you live now.

Look carefully at **Figure 18-12**; it summarizes the entire history of the Universe, from the sorting out of stable particles in



NASA/JPL-Caltech/GSFC

▲ **Figure 18-11** A map of the far-infrared background of a region in the constellation Ursa Major made using the *Spitzer Space Telescope*. Foreground stars, galaxies, and other known sources have been removed by computer processing. The remaining background glow is understood to be from objects that formed in the first billion years of the Universe's history, such as massive stars and voracious black holes at the centers of forming galaxies. Yellow and white indicate the brightest parts of the background.

Science: A System of Knowledge

What is the difference between believing in the Big Bang and understanding it?

If you ask a scientist, "Do you believe in the Big Bang?" she or he may hesitate before responding. The question implies something incorrect about the way science works.

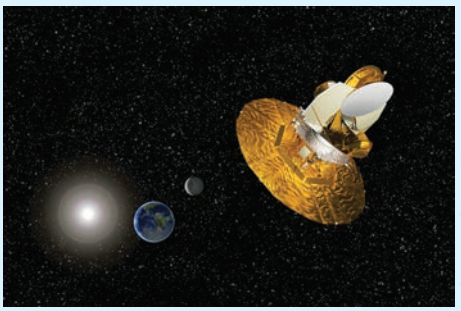
The goal of science is to understand nature. Science is a logical process based on observations and experiments used to test and confirm hypotheses and theories. A scientist does not really believe in even a well-confirmed theory in the way people normally use the word *believe*. Rather, the scientist understands the theory and recognizes how different pieces of evidence support or contradict the theory.

There are other ways to know things, and there are many systems of belief. Religions, for example, are systems of belief that are not entirely based on observation. In some cases, a political system is also a system of belief; many people believe that democracy is the best form of government and do not ask for, or expect, evidence supporting that belief. A system of belief can be powerful and lead to deep insights, but it is different from science.

Scientists try to be careful with words, so thoughtful scientists would not say they *believe* in the Big Bang. They would say that the evidence is overwhelming that the Big Bang really did occur and that they are compelled by a logical analysis of both the observations and the theory to conclude that the theory is very likely correct. In this way scientists try to be objective and reason without distortion by personal feelings and prejudices.

A scientist once referred to "the terrible rule of evidence." Sometimes the evidence forces a scientist to a conclusion she or he does not like, but the personal preferences of each scientist must take second place to the rule of evidence.

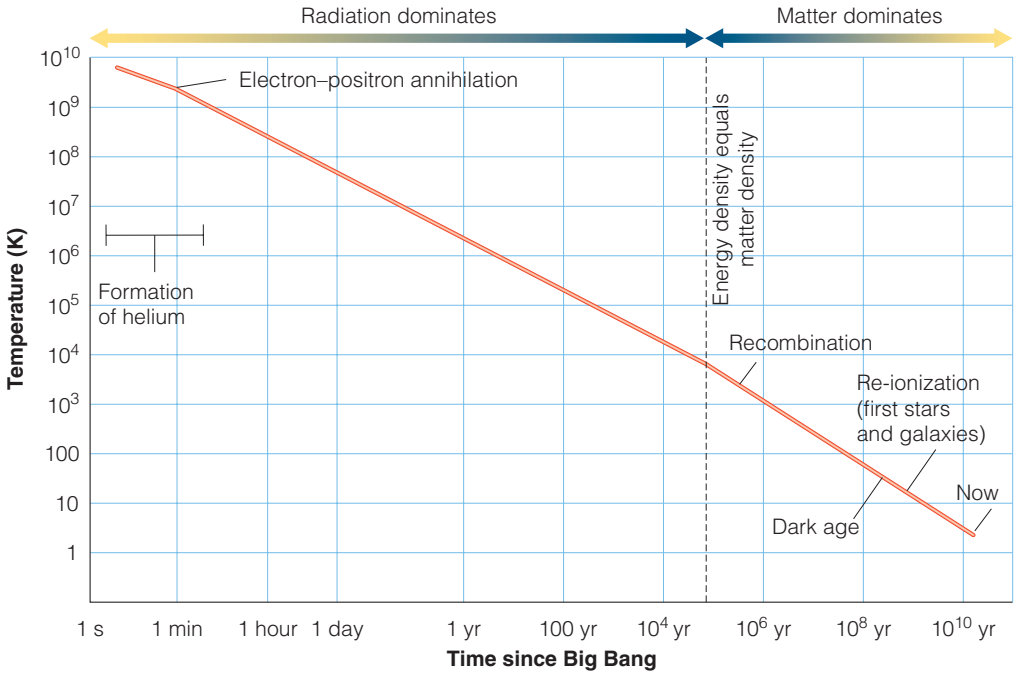
Do you believe in the Big Bang? Or, instead, do you have confidence that the theory is right because of the evidence? There is a big difference.



NASA/*WMAP* Science Team

Scientific knowledge is objective, based on evidence such as that gathered by spacecraft.

the earliest moments after the Big Bang, through formation of helium in the first 3 minutes, to the eras of energy–matter equality and recombination, formation of the first stars, and re-ionization of the gas, and finally up to the present day. It may seem amazing that mere humans limited to Earth can draw such a diagram, but remember that it is based on evidence and on the best understanding of how matter and energy interact (**How Do We Know? 18-2**).



◀ Figure 18-12 During the first few minutes of the Big Bang, some hydrogen fused to produce helium, but the Universe quickly became too cool for such fusion reactions to continue. The rate of cooling increased as matter began to dominate over radiation. Recombination freed the radiation from the influence of the gas, and re-ionization was caused by the birth of the first stars. Note how the exponential scale in time stretches early history and compresses recent history.

Practicing SCIENCE

How does the cosmic microwave background radiation provide evidence that there was a Big Bang? If there was a Big Bang, then the matter of the Universe was originally in a hot, dense, opaque state that would have produced blackbody radiation. Once the expansion of the Universe reached the stage where the matter became transparent, the radiation trapped between matter particles would be free to travel through space. Therefore, there should be background blackbody radiation coming from all directions, redshifted to microwave wavelengths by subsequent expansion of the Universe.

Evidence confirms that the cosmic microwave background radiation has exactly the expected characteristics: It comes with equal intensity from all directions, its source lies "behind" all other types of emission, and it has the spectral energy distribution of a blackbody. When we detect those photons, we are, in a sense, "seeing" the Big Bang.

Observation of the cosmic microwave background radiation is powerful evidence that there was a Big Bang. That tells you how the Universe began, but you should examine a point that the original theoretical prediction assumed: ***Why is it true that the Universe cannot have an edge or a center?***

18-3 Space and Time, Matter and Gravity

How can the Big Bang have happened everywhere? How did that moment of astonishing heat and density lead to the present-day Universe of stars and galaxies? To investigate these questions, once again you need to put seemingly reasonable expectations aside and let the observations plus a little bit of careful reasoning take you wherever they lead.

18-3a The View from Here

The Universe is **isotropic**, meaning it looks about the same in whichever direction you observe. Of course, if you look toward a galaxy cluster, you see more galaxies than you see in other directions, but that is only a local variation. On the average, you see similar numbers of galaxies in every direction. Furthermore, the background radiation is almost perfectly uniform across the sky. The Universe is observed to be very isotropic.

The Universe is also observed to be **homogeneous**, which means it is about the same at all locations. Of course, there are local variations; some regions contain more galaxies and some fewer. Also, because the Universe evolves, at large look-back times you can see galaxies at an earlier developmental stage. But, if you account for those well-understood variations, then the Universe seems to be, on average, the same everywhere.

The Universe is evidently both isotropic and homogeneous. (Those two properties don't necessarily go together; for an intellectual exercise, try to imagine arrangements that are isotropic but not homogeneous, or homogeneous but not isotropic.) The fact that the Universe is both isotropic and homogeneous leads to the **cosmological principle**: *Any observer in any galaxy sees the same general properties for the Universe*. (Again, this overall principle intentionally overlooks minor local variations.)

The cosmological principle implies there are no special places in the Universe. What you see from the Milky Way Galaxy is typical of what all intelligent creatures see from their respective home galaxies. Furthermore, the cosmological principle states, in a new form, the idea that the Universe can have no center or edge. A center or an edge would be a special place, and the cosmological principle says there are no special places.

18-3b Cosmic Redshifts

Einstein's theory of general relativity, published in 1916, describes the fabric of the Universe, space and time considered together, as one entity: space-time. That idea will give you a new insight into what it means to say the Universe is expanding.

General relativity describes space-time as if it is made of stretching rubber, and that explains one of the most important observations in cosmology—redshifts. It is a **Common Misconception**, even among scientists, that cosmological redshifts are Doppler shifts of galaxies flying away through space. Instead, except for relatively small, local motions within clusters of galaxies, galaxies are all approximately at rest and have always been so. The galaxies are being *separated* from each other as space-time expands. Also, as space-time expands it stretches any photon traveling through space so that the photon's wavelength increases. Photons from distant galaxies spend more time traveling through space and are stretched more—in other words, have larger redshifts—than photons from nearby galaxies. That is why redshift depends on distance.

Astronomers often express redshifts as if they were actual radial velocities, but the redshifts of the galaxies are not Doppler shifts. That is why we are careful to refer to a galaxy's *apparent* velocity of recession. You can even find some textbooks that convert large cosmological redshifts to velocities near the speed of light using Einstein's relativistic Doppler equation, but that formula applies to motion through space and not to the behavior of space-time itself. So, the relativistic Doppler formula should not be used in a cosmological context. Nevertheless, the Hubble law still applies: Redshifts indicate distances to galaxies because redshifts show how much the Universe has expanded, and thus how much time has elapsed, since the photons from those galaxies started on their journey.

18-3c The Shape of Space

Almost immediately after Einstein published his general relativity theory, other theorists were able to solve the sophisticated mathematics and compute simplified descriptions of the

behavior of space-time and matter. The resulting model universes dominated cosmology throughout the 20th century.

The general relativity equations allowed three possibilities for space-time as a whole: It might be curved in either of two different ways, or it might have no curvature at all. Most people find these curved models difficult to imagine, and modern observations have shown that the simplest model, without curvature, is almost certainly correct; so, you may not need to wrap your brain around the curved models. You might, however, like to know a few of their most important properties.

Study **The Nature of Space-Time**, and notice three important points and three new terms:

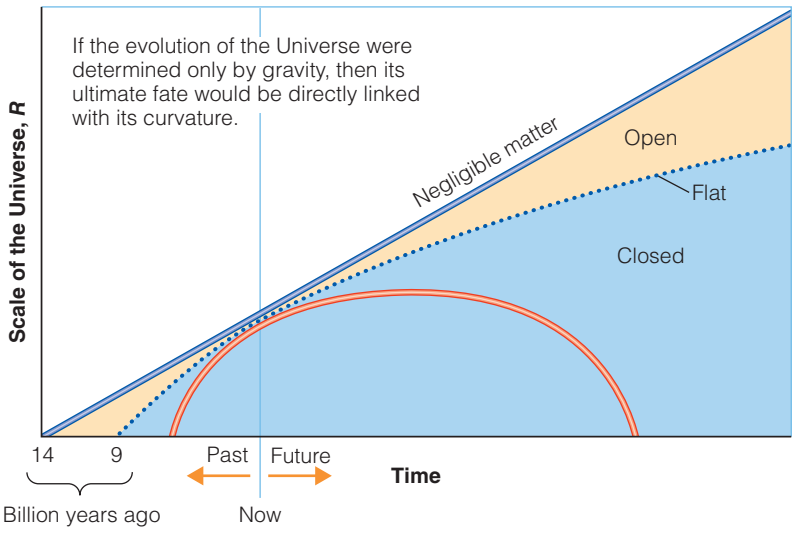
1. Einstein described space-time as if it were a rubber canvas on which the Universe is painted. The gradual stretching of space-time lengthens the wavelengths of photons on their way from distant galaxies to Earth. Redshifts of galaxies are a result of this stretching of the canvas, not Doppler shifts.
2. Space-time could be curved. You would not notice this curvature in daily life; it would only be evident in measurements involving very distant objects. A model universe with positive curvature would be finite, referred to as a *closed universe*. The model with zero curvature is called a *flat universe*. A flat universe model would be infinite, which is termed an *open universe*. In this model the rules of geometry across cosmological distances would be the same as the rules of Euclidean geometry you learned in high school. A model universe with negative curvature is also open and infinite. Note that the Universe would not have an edge or a center in any of these models, closed or open.
3. For most of the 20th century, cosmologists attempted to measure or infer the amount of curvature of space-time. Modern observations indicate that the geometry of the Universe is almost certainly flat (uncurved), and therefore the Universe is probably infinite.

18-3d Gravity and the Evolution of the Universe

Throughout the 20th century, observations could not eliminate closed models, so they were considered a real possibility. Most closed universe models predicted that the expansion of the Universe would eventually become a contraction that would bring all of the matter and energy back to the high-density Big Bang state, which was sometimes called the "big crunch." In contrast, the Universe expands forever in both the flat and negatively curved models.

Cosmologists struggled to find observations or logical arguments that would allow them to choose among these three kinds of models. The main criterion is density. According to general relativity, the overall curvature of space-time is determined by the average density of matter plus energy in the Universe. Using an updated value of the Hubble constant, cosmologists calculate that space-time is flat if the average density of the Universe equals a **critical density** of 9×10^{-27} kg/m^3. (Note that this is equal to about five protons per cubic meter.) If the average density of the Universe is more than the critical density, the Universe must be closed; if it is less, the Universe must be open. Attempts to measure the actual value of the Universe's density, however, were difficult and inconclusive.

The "shapes" of the past and future for these three kinds of universe models are illustrated in **Figure 18-13**, which compares their respective histories of expansion versus time. The parameter R on the vertical axis is a measure of the extent to which the Universe has expanded. Roughly speaking, R can be thought of as the average distance between galaxies. You can see in the diagram that closed universes expand and then contract, whereas both



◀ **Figure 18-13** Illustrations of some simple models of the Universe that depend only on the effect of gravity, showing their respective histories of Universe scale factor R versus time. Open universe models expand without end, and the corresponding curves fall in the region shaded orange. Closed models expand and then contract again (*red curve*). A flat universe (*dotted line*) marks the boundary between open and closed universe models. The age of the Universe for each model is equal to the horizontal distance from "Now" back to the time when R was zero. A model universe with negligible matter would have expanded at a constant rate (*blue line*) and have an age of about 14 billion years for a Hubble constant value of 70 km/s/Mpc (Eq. 18-1).

flat and open universes expand forever. Note that Figure 18-13 describes the evolution of universe models that depend only on gravity; in a later section you will learn how the history of the Universe changes if there is something that can counteract the cosmological effects of gravity.

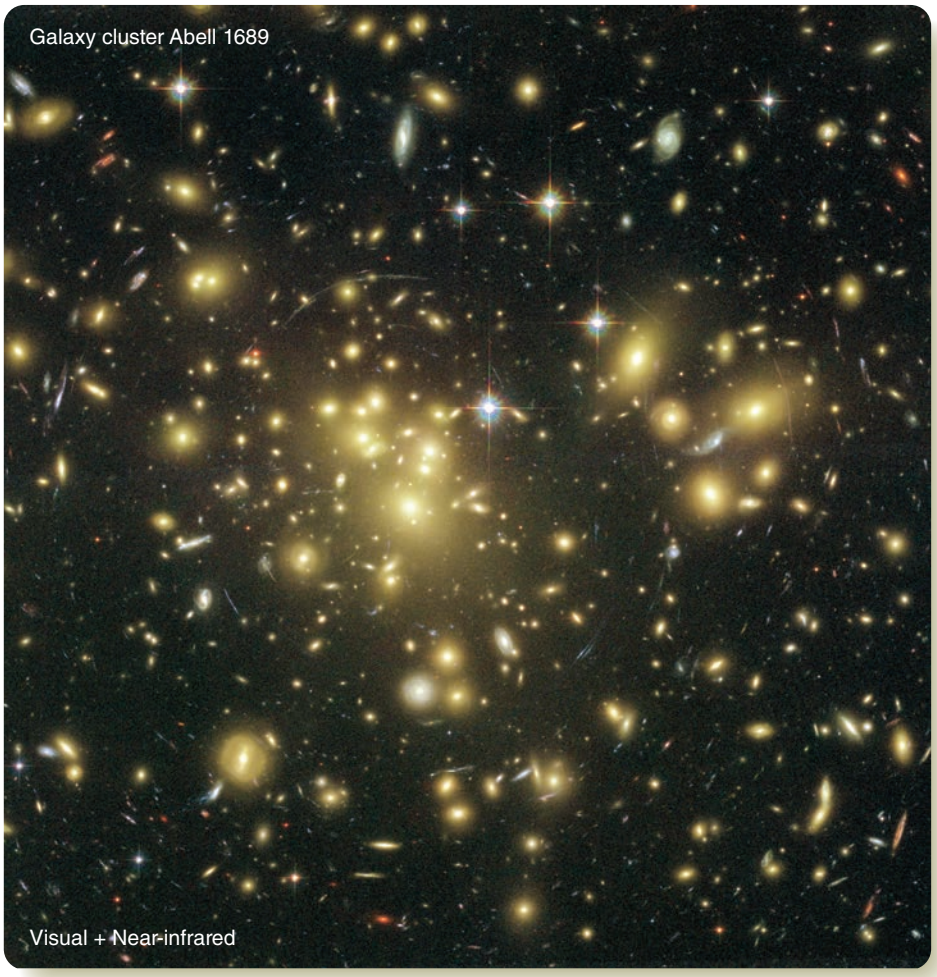
You can see in Figure 18-13 that the Hubble law can't be used to find the actual age of the Universe unless you know whether it is open, closed, or flat. In a previous section of this chapter you calculated the Hubble time, an estimate of the age of the Universe, as the reciprocal of the Hubble constant H_0 that expresses the Universe's observed expansion rate. The Hubble time is the age the Universe would have if it were totally open, meaning it contains negligible amounts of matter and has nearly zero density. Instead, if the Universe is actually dense enough to have flat geometry, then its true age would be two-thirds of the Hubble time.

Hubble Space Telescope measurements made it clear that the Hubble constant is close to 70 km/s/Mpc, corresponding to a Hubble time of 14 billion years. But as you will learn later in this chapter, theorists argued that the Universe must be flat. If the Universe is flat and H_0 is 70 km/s/Mpc, its true age should be only about 9 billion years. However, most globular star clusters are known to be significantly older than that from observations of their H–R diagram turnoff points. Putting all these clues together, some astronomers speculated that the Universe must be much less dense than the critical density, negatively curved, open, and infinite. But, is that correct?

18-3e Ordinary Matter and Dark Matter

The total amount of matter, ordinary matter plus dark matter, is a crucial piece of the puzzle regarding the density, geometry, age, and future fate of the Universe. You know that there is plenty of dark matter along with ordinary matter in galaxies and galaxy clusters (Sections 15-2b and 16-2f). But is the total of both enough to reach the critical density and make the Universe flat, or even more, which would make the Universe positively curved and closed? Or, is the total density less than the critical density and the Universe is open and negatively curved?

Dark matter is obviously an important component of the Universe. Look at **Figure 18-14**, dramatic evidence of dark matter revealing itself by gravitational lensing. The galaxy cluster in the



NASA/ESA/STScI/AURA/NSF, N. Benitez (JHU), T. Broadhurst (RIP, HUJI), H. Ford (JHU), M. Clampin (STScI), G.Hartig (STScI), G. Illingworth (UCO/Lick Observatory), ACS Science Team

◀ **Figure 18-14** Gravitational lensing shows that galaxy clusters contain much more mass than what is visible. The yellowish galaxies in this image are members of a relatively nearby cluster of galaxies. Most of the objects in this image are blue or red images of very distant galaxies focused by the gravitational field of the cluster. Some of these imaged galaxies may be more than 13 billion ly away. That they can be seen at all is evidence that the foreground galaxy cluster contains large amounts of dark matter.

The Nature of Space-Time

1 In 1929, Edwin Hubble discovered that the redshifts of the galaxies are proportional to distance—a relationship now known as the Hubble law. It was immediately understood to be compelling evidence that the Universe is expanding.

Distance is the separation between two points in space, and interval is the separation between two moments in time. Einstein's theories of special and general relativity show how distances in space and intervals in time are interrelated, so that space and time should be considered together as space-time. You can think of space-time as a canvas on which the Universe is painted, a canvas that can stretch.

NASA/ESA/STScI/AURA/NSF/*Hubble* Heritage Team/
G. Meurer, T. Heckman, & M. Sirianni (JHU), and C. Leitherer,
J. Harris, & D. Calzetti (STScI)

1a Astronomers often express galaxy redshifts as apparent velocities of recession, but these redshifts are not Doppler shifts. They are caused by the expansion of space-time.

For decades textbooks have described the cosmological redshifts using Einstein's relativistic Doppler formula, but that formula applies to motion through space and not to the behavior of space itself.

Galaxies are not moving through space as the Universe expands any more than the raisins in the raisin bread analogy are swimming through the dough as the bread rises (Figure 17-5). Except for small relative orbital motions among neighbors, galaxies are motionless in space-time and are being carried away from each other by the stretching of space-time.

1b The stretching of space-time not only moves the galaxies away from each other, but also increases the wavelengths of photons traveling through space-time. See the illustration below.

What are the cosmological redshifts?

A distant galaxy emits a short-wavelength photon toward our Galaxy.

Grid shows expansion of space-time.

The expansion of space-time stretches the photon to longer wavelength as it travels.

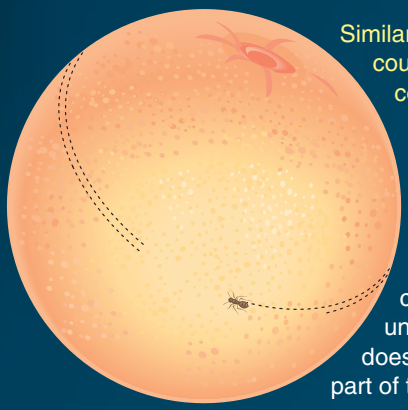
The farther the photon has to travel, the more it is stretched.

When the photon arrives at our Galaxy, we see it with a longer wavelength—a redshift that is proportional to the distance traveled.

Earth

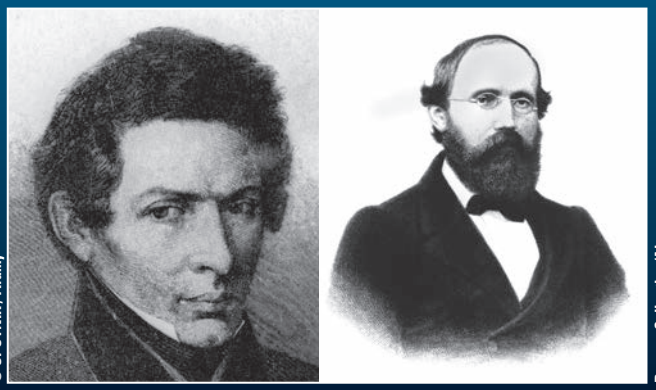
1c A photon traveling to Earth from a more distant galaxy spends more time moving through space, and is therefore stretched more, than a photon from a nearby galaxy. That is why cosmological redshifts are proportional to distance.

2 To think about space-time curvature, you can use an analogy of a two-dimensional ant on an orange. If the ant is truly two-dimensional, it can travel only forward and back, right and left; it will not perceive up and down. As it walks around this mini-universe, it doesn't realize this is actually the surface of a three-dimensional sphere and will eventually realize it has been everywhere because its universe is covered by its footprints. The ant will conclude that it lives in a finite universe that has no edge and no center.



Similarly, a three-dimensional universe could be curved back on itself so that it could be finite, but an explorer would never find an edge. With no edge, such a universe would also have no center.

Notice that the center of the three-dimensional orange is not the center of the ant's two-dimensional universe—the center of the orange does not lie on the surface and is not part of the ant's universe.



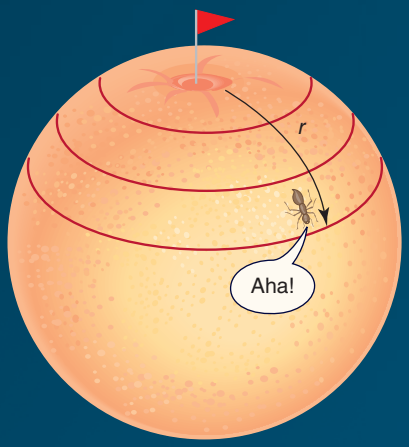
© SPUTNIK / Alamy

Nikolai Lobachevsky
1792–1856

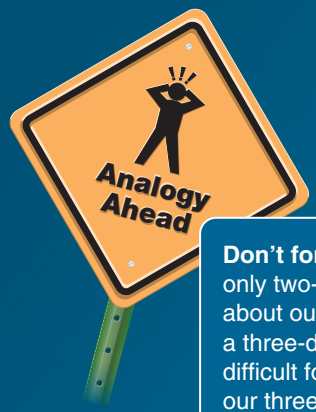
Bernhard Riemann
1826–1866

Bygone Collection/Alamy

2a During the 19th century, two mathematicians, the Russian Nikolai Lobachevsky and the German Bernhard Riemann, developed the mathematics of curved surfaces—non-Euclidian geometry. That, with Einstein's theory of general relativity, showed that space-time might be curved by an amount that depends on the average density of the Universe. Curved space-time dominated 20th-century cosmology.

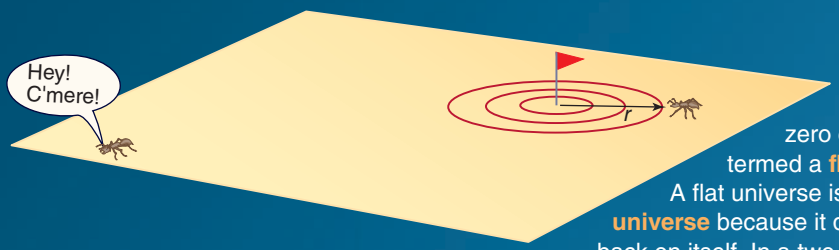


There are three possible types of curvature for a model universe: positive, zero, or negative. A positively curved universe can be represented by a sphere, a zero curvature universe by a flat surface, and a negatively curved universe by a saddle shape.



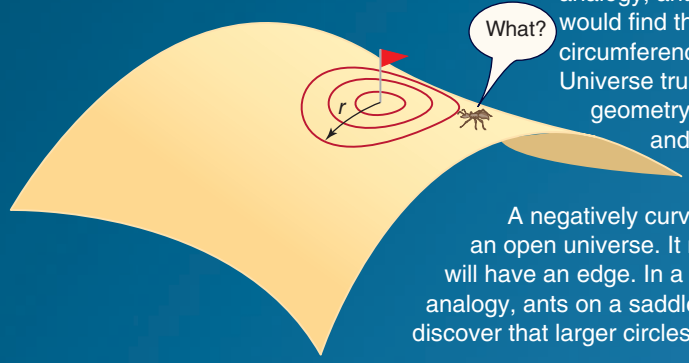
Don't forget: These analogies are only two-dimensional. When you think about our Universe, you must think of a three-dimensional universe. It is as difficult for you to imagine curvature of our three-dimensional universe as it is for ants to imagine curvature of their two-dimensional universes.

A positively curved model universe is termed a **closed universe** because it curves back on itself. It is finite, but it has no edge. In a two-dimensional analogy, ants on an orange could discover that they are in a curved universe because the circumferences of large circles are less than $2\pi r$.



A model universe with zero curvature is termed a **flat universe**. A flat universe is also an **open universe** because it does not curve back on itself. In a two-dimensional analogy, ants on a flat sheet of paper would find that all circles have circumferences equal to $2\pi r$. If the Universe truly has flat (Euclidean) geometry, then it must be infinite and has no center.

3 Throughout the 20th century, cosmologists tried to determine whether our Universe is actually positively curved, negatively curved, or uncurved (flat) by making difficult and inconclusive measurements analogous to the ants measuring the circumferences of large circles. As you will see later in this chapter, new ways to measure the curvature show that the Universe is close to exactly flat.



A negatively curved model is also an open universe. It must be infinite or it will have an edge. In a two-dimensional analogy, ants on a saddle shape will discover that larger circles have

figure contains so much dark matter that it warps space-time and focuses the images of more distant galaxies into short arcs of light. A rough estimate based on observations of galaxies and galaxy clusters is that dark matter outweighs ordinary matter by a factor of between 5 and 10. Looking at the universe of visible matter in galaxies is like looking at a tree and seeing only the leaves.

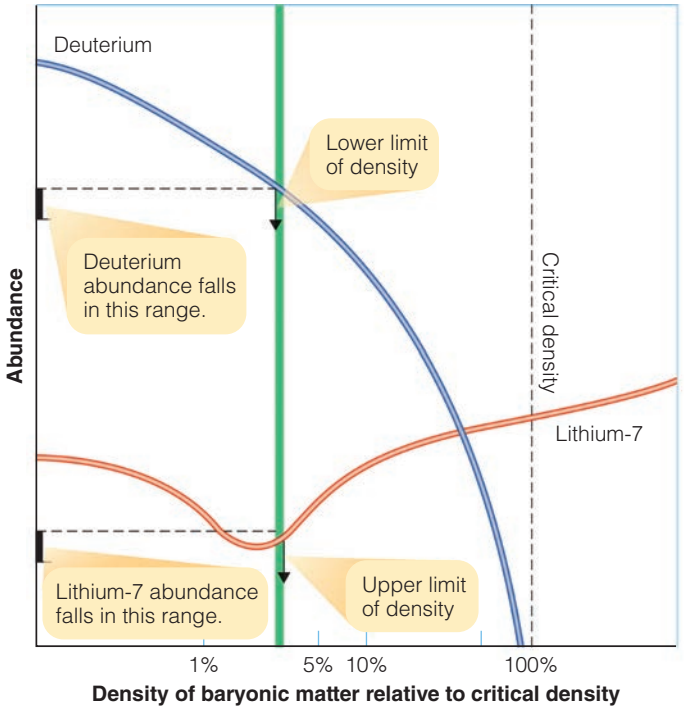
One way to make progress on estimating the total density of ordinary matter plus dark matter would be to determine the density of ordinary matter. You can do that by starting with what is known about conditions right after the Big Bang. As you have learned, during the first few minutes of the Universe's history, nuclear reactions converted some protons into helium and a very small amount into other elements. The amount of elements heavier than helium that could be produced during the Big Bang were controlled by the density of ordinary matter. For example, if the density of matter had been relatively high so there were lots of ordinary matter particles such as protons and neutrons flying around, they would have collided with deuterium (hydrogen-2) nuclei and converted most of them into helium. On the other hand, if the density of ordinary matter particles had been low, more deuterium would have survived intact without being converted to helium.

Figure 18-9 shows that there is a gap between helium and lithium; there is no stable nucleus with atomic mass 5, so regular nuclear reactions during the first few minutes of the Big Bang had difficulty converting helium into lithium. Only if the density of protons and neutrons was high enough could a few nuclear reactions have "leaped the gap" and produced a small amount of the isotope lithium-7.

As shown in **Figure 18-15**, if you can measure the amount of deuterium in the Universe after the Big Bang finished, it gives you a lower limit on the density of the Universe relative to the critical density, and the abundance of lithium-7 gives you an upper limit. To complicate matters, both deuterium and lithium are destroyed by nuclear reactions in stars, so astronomers have attempted the difficult task of identifying and studying the composition of clouds of gas at large look-back times that had not yet been altered by nuclear reactions in stars.

The surprising conclusion from these observations and calculations is that ordinary matter from which you, Earth, and the stars are made make up only 4 to 5 percent of the critical density. This means that, based on observations of the ratio of ordinary matter to dark matter in galaxies and galaxy clusters, dark matter plus ordinary matter must be less than about 50 percent of the critical density. If that's all there is, it's clearly not enough to make the Universe flat or closed.

The protons and neutrons that make up normal matter belong to a family of subatomic particles called *baryons*, so cosmologists believe that most of the dark matter cannot be baryons. Only a few percent of the mass in the Universe can be ordinary baryonic matter; the dark matter must be **nonbaryonic matter**.



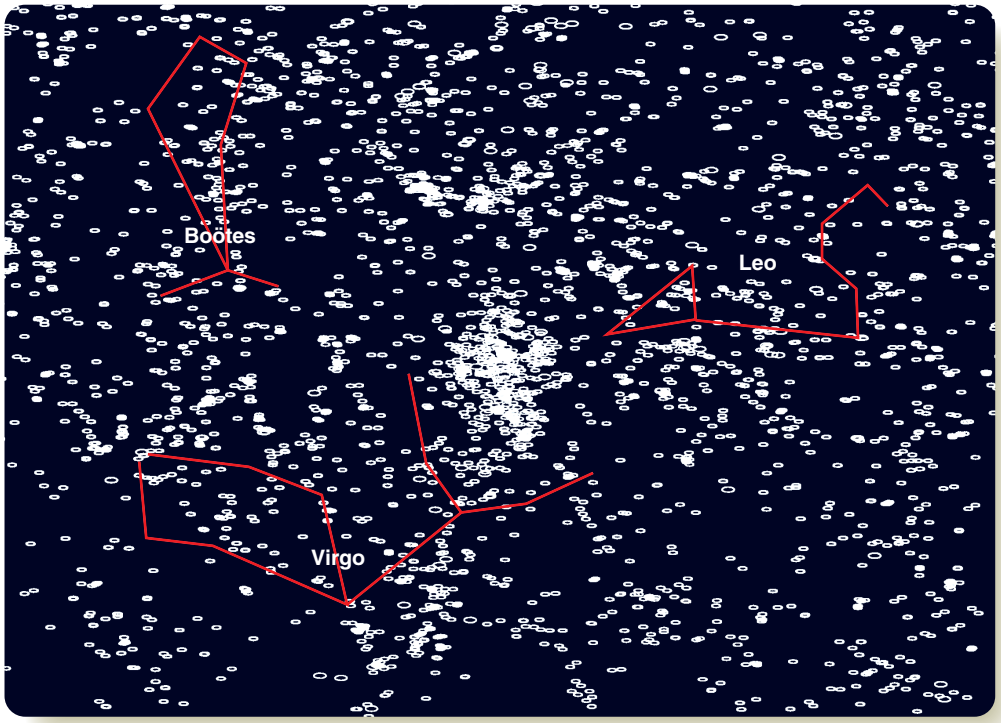
▲ **Figure 18-15** This diagram compares observation with theory. Theory predicts how much deuterium (*blue curve*) and lithium-7 (*red curve*) you would observe for different densities of normal matter. The observed density of deuterium falls in a narrow range shown at upper left and sets a lower limit on the possible density of normal matter. The observed density of lithium-7, shown at lower left, sets an upper limit. This means the true density of normal matter must fall in a narrow range represented by the green column. Certainly, the density of normal matter is much less than the critical density.

Some theorists thought that neutrinos, particles that are predicted to be very abundant in the Universe yet are not baryons, might have enough mass to make up the dark matter, but later measurements show that neutrinos are not massive enough. Some particle physics theories predict the existence of new type of particles labeled **WIMPs** (weakly interacting massive particles), but WIMPs have not been detected with certainty in the laboratory. The true nature of dark matter remains one of the major mysteries of astronomy, but its effects are seen everywhere.

18-3f Origin of Large-Scale Structure

As you have already learned, the Universe seems homogeneous on the largest scales. That is, each location is, on average, pretty much like any other location. But on smaller scales, there are irregularities. The sky is filled with galaxies and clusters of galaxies, and there are even bigger aggregations that astronomers call **large-scale structure**. Studies of large-scale structure lead to insights about how the Universe has evolved.

Galaxies are observed in clusters ranging from a few galaxies to thousands, and those clusters appear to be grouped into

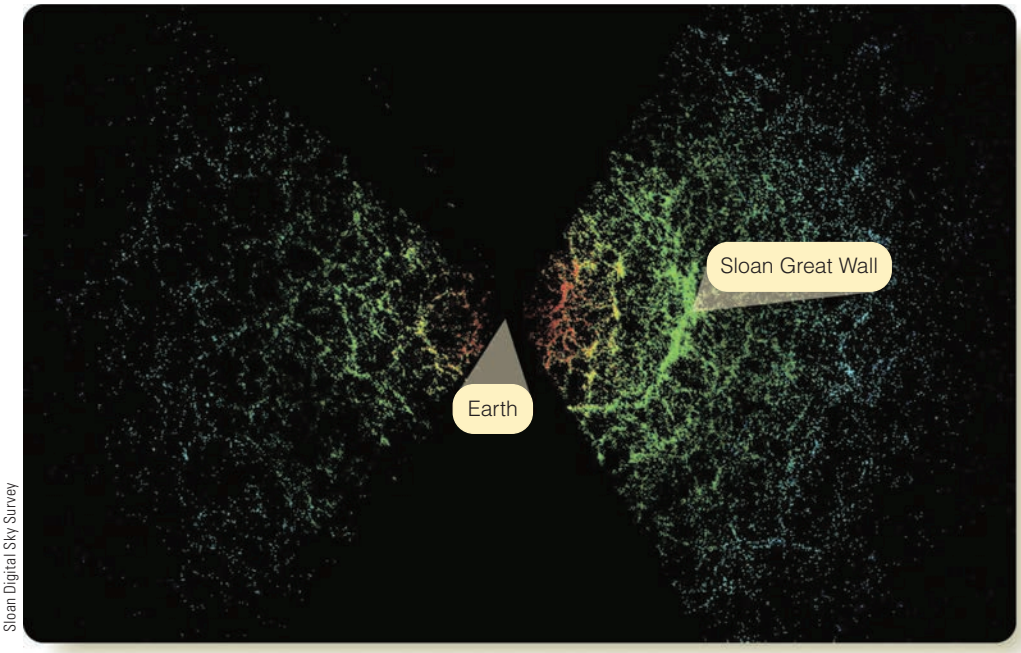


Michael A. Seeds/data from Carina Software & Instruments, Inc; © 2016 Cengage

◀ **Figure 18-16** The distribution of brighter galaxies in the sky reveals the great Virgo cluster (*center*), containing more than 1000 galaxies only about 17 Mpc away. Other clusters fill the sky, such as the more distant Coma cluster just above the Virgo cluster in this diagram. The Virgo cluster is linked with others to form the Local Supercluster.

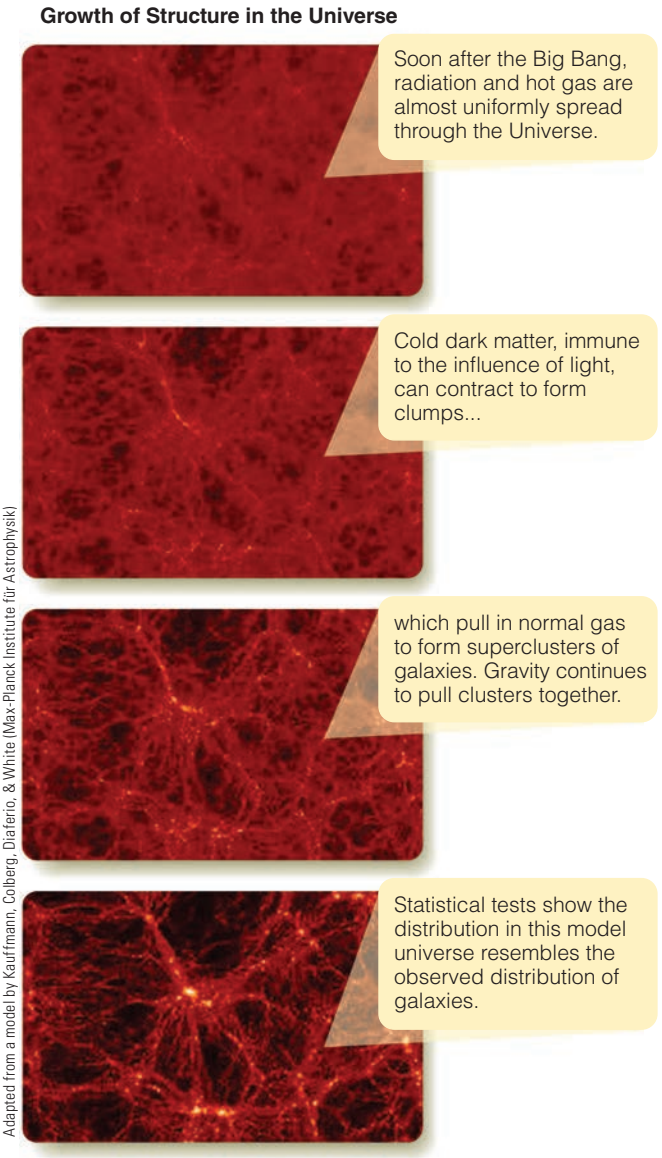
superclusters (Figure 18-16). The Local Supercluster, in which we live, is a roughly disk-shaped swarm of galaxies and galaxy clusters 50 to 75 Mpc in diameter. By measuring the redshifts and positions of hundreds of thousands of galaxies in great slices across the sky, astronomers have been able to create maps revealing that superclusters are not scattered at random. They are distributed in long, narrow **filaments** of galaxies and thin walls that outline great **voids** that are nearly empty of galaxies (Figure 18-17).

The observed large-scale structure is puzzling because the cosmic microwave background radiation is very smooth, which means that the density of the Universe must have been extremely uniform at the time of recombination, about 400,000 years after the Big Bang. Yet the farthest galaxies and quasars known are seen less than 1 billion years after the Big Bang. How did matter that was so uniform at the time of recombination become lumpy and condense in that amount of time to form galaxies? In other



Sloan Digital Sky Survey

◀ **Figure 18-17** Nearly 70,000 galaxies are plotted in this double slice of the Universe extending outward in the plane of Earth's equator. The nearest galaxies are shown in red and the more distant in green and blue. The galaxies form filaments and walls enclosing nearly empty voids. The Sloan Great Wall, discovered by the Sloan Digital Sky Survey (Section 6-3a), is almost 1.4 billion ly long and is the largest known structure in the Universe. The most distant galaxies in this diagram are roughly 3 billion ly from Earth.


Growth of Structure in the Universe
Soon after the Big Bang, radiation and hot gas are almost uniformly spread through the Universe.
Cold dark matter, immune to the influence of light, can contract to form clumps...
which pull in normal gas to form superclusters of galaxies. Gravity continues to pull clusters together.
Statistical tests show the distribution in this model universe resembles the observed distribution of galaxies.
Adapted from a model by Kauffmann, Colberg, Diaferio, & White (Max-Planck Institute für Astrophysik)

▲**Figure 18-18** This computer model traces the formation of structure in the Universe from soon after the Big Bang to the present.

words, how did the galaxies, clusters of galaxies, and supermassive black holes we observe as quasars (Section 17-1c) form so early in the history of the Universe?

The answer does not lie with baryonic (ordinary) matter. Baryonic matter is so rare in the Universe that cosmologists can calculate that it did not have enough gravity to pull itself together quickly after the Big Bang, fighting against the tendency of the strong radiation pressure in the early Universe to smooth out any density variations. So long as radiation dominated the Universe, it prevented clouds of ordinary matter from contracting to begin forming galaxies. In contrast, dark matter is dark because it does not interact with light, so it was immune to the intense radiation and could condense into clumps while the Universe was very young. Dark matter could have given formation of galaxies a head start soon after the Big Bang. Once the Universe became dominated by matter instead of by radiation, the baryonic matter could begin falling into the waiting pockets of dark matter to form galaxies and larger structures.

In fact, details of the observed large-scale structure allow theorists to choose between different hypothetical types of dark matter. One dark matter candidate called **hot dark matter** would be composed of particles moving at, or near, the speed of light. Such fast-moving particles would not clump together easily and could not have stimulated the formation of objects as small as galaxies or even clusters of galaxies. Instead, models with **cold dark matter**, composed of particles moving much slower than light that are able to clump into relatively small structures, are most successful at predicting the formation of galaxies and clusters of galaxies with the right sizes and in the right amount of time to match the observations (**Figure 18-18**).

But what started the clumping of the dark matter? Theorists say that space is filled with tiny, random quantum mechanical fluctuations of energy, like bubbles smaller than the smallest atomic particles, forming and vanishing continuously. At the moment of the Big Bang, those fluctuations would have been present as variations in the density and temperature of the Universe—the laws of quantum mechanics require that the Universe could not have been completely smooth. As the Universe expanded, those tiny fluctuations would have been stretched to very large but very subtle variations in gravitational fields that could have stimulated the formation of galaxy superclusters, filaments, and walls. The structure you see in Figures 18-16 and 18-17 may be the present-day ghostly traces of microscopic random fluctuations in the infant Universe.

Practicing **SCIENCE**

Why do cosmologists think that dark matter can't be baryonic? The evidence is very compelling, but scientists need a theoretical framework to interpret evidence, in this case a bit of nuclear physics.

Cosmologists can calculate that small amounts of isotopes like deuterium and lithium-7 were produced by nuclear reactions in the first minutes of the Big Bang. The resulting abundance of those elements depends strongly on how many protons and neutrons were available at that time. Because those particles belong to the family of particles called *baryons*, physicists refer to normal matter as baryonic. Measurements of the abundances of deuterium and lithium-7 show that the Universe cannot contain more baryons than about 5 percent of the critical density. Yet observations of galaxies and galaxy clusters show that dark matter must make up almost 30 percent of the critical density. Consequently, cosmologists conclude that the dark matter must be made up of nonbaryonic particles.

Finding the dark matter is important because the density of matter in the Universe determines the curvature of space-time. Now consider a related question. ***How does the modern understanding of space-time explain cosmic redshifts?***

18-4 21st-Century Cosmology

If you are a little dizzy from the weirdness of expanding spacetime, dark matter, and quantum fluctuations, make sure you are sitting down before you read further. As the 21st century began, astronomers made a discovery that startled all cosmologists: The expansion of the Universe is actually accelerating. A few years later, another group of astronomers made measurements that confirmed a hypothesis called "inflation," which proposes that the Universe went through a brief moment of superexpansion when it was a tiny fraction of a second old. You'll have to go back a few decades to understand how those two amazing new discoveries fit well with some of the things you have learned already in this chapter.

18-4a Acceleration and Dark Energy

Both common sense and the theory of general relativity suggest that as the galaxies recede from each other, the expansion should be opposed and slowed by gravity trying to pull the galaxies toward each other. How much the expansion is slowed should depend on the amount of matter in the Universe. If the density of matter and energy in the Universe is less than the critical density, the expansion should be slowing only moderately, and the Universe would be expected to expand forever. If the density of matter and energy is greater than the critical density, the expansion should be slowing down now significantly, and the Universe should eventually begin contracting.

For decades, astronomers struggled to measure carefully the redshifts and distances of large samples of galaxies in order to detect the slowing of the expansion. You can understand that detecting not just the expansion, but a change in the rate of expansion, would be quite difficult because it requires accurate measurements of the distances to very remote galaxies.

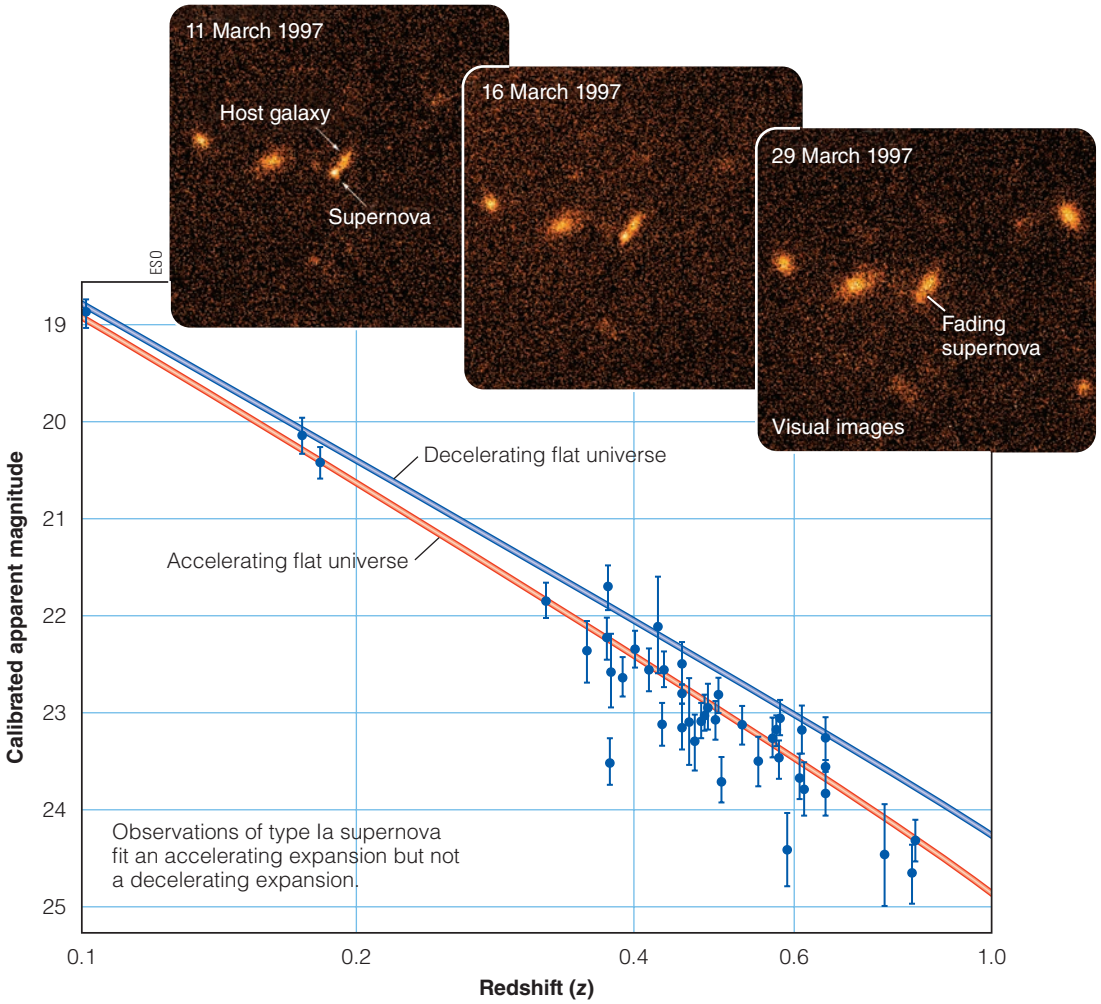
After its launch in 1990, the *Hubble Space Telescope* made it possible to measure distances to galaxies with unprecedented accuracy, and two competing research teams began using the same technique of calibrating type Ia supernovae as distance indicators. As you learned in Section 13-4d, a type Ia supernova occurs when a white dwarf gains matter from a companion star, exceeds the Chandrasekhar limit, and collapses in a supernova explosion. Because all such white dwarfs should collapse at the same mass threshold, they should all produce explosions of the same size and luminosity, which makes them good distance indicators ("standard bombs"). The two teams calibrated type Ia supernovae by locating such supernovae occurring in nearby galaxies that had distances known from Cepheid variables and other reliable distance indicators. Once the peak luminosity of type Ia supernovae had been determined, they could be used to find the distances of much more distant galaxies.

Both SN Ia cosmology teams announced their results in 1998, in agreement that the expansion of the Universe is not slowing down. Contrary to expectations, it is speeding up! That is, the expansion of the Universe is accelerating (**Figure 18-19**). For their work, leaders from both teams, Saul Perlmutter, Brian Schmidt, and Adam Reiss, shared the 2011 Nobel Prize in Physics.

The claim that the expansion of the Universe is accelerating was totally unexpected, and astronomers all over the world immediately began checking it. This result depends critically on the calibration of type Ia supernovae as distance indicators, and some astronomers suggested that the calibration might be wrong (look back to How Do We Know? 15-1). However, problems with the calibration have been ruled out by subsequent observations of supernovae at even greater distances by the original two teams plus other research groups. Also, the Two-Degree-Field Redshift Survey mapped the position and redshift of 250,000 galaxies and 30,000 quasars. As expected, the galaxies are spread in filaments and walls, and a statistical analysis of the distribution of galaxies also shows that the Universe expansion is accelerating. This is an important result because it is a confirmation of acceleration that is independent of the brightness of type Ia supernovae When a hypothesis is confirmed by observations of several different types, scientists have much more confidence that it is a true description of nature. The Universe really does seem to be expanding faster and faster rather than slowing down.

If the expansion of the Universe is accelerating, then there must be a force of repulsion in the Universe that counteracts gravity, and cosmologists are struggling to understand what it could be. One possibility was suggested by Einstein in 1916; he recognized that the equations of general relativity implied that the Universe should not be static. Galaxies would not stay at constant distances from each other because their gravity should make the scale of the Universe decrease. The only solutions seemed to be either a universe that is contracting under the influence of gravity or a universe in which the galaxies are rushing away from each other so rapidly that gravity cannot pull them together. Neither possibility seemed reasonable. In 1916, cosmologists did not yet know that the Universe is expanding, so Einstein made a change to his theory that he later thought was a mistake.

To balance the attractive force of gravity, Einstein added a constant to his equations called the **cosmological constant**, represented by an uppercase lambda (Λ). That constant represents a force of repulsion that balances the gravitational attraction between galaxies so the Universe won't contract or expand. Thirteen years later, Hubble announced his observations indicating that the Universe is not static but expanding. Einstein then famously said that introducing the cosmological constant "fudge factor" to make the Universe static was the biggest blunder of his career. Modern cosmologists think he may have been right after all.



◀ **Figure 18-19** From the way supernovae fade over time, astronomers can identify those that are type Ia. Once calibrated, the peak brightness of each of those supernovae could be compared with their respective redshifts, revealing that type Ia supernovae in the range of redshifts plotted were about 25 percent fainter than expected. That must mean they are farther away than expected, given their redshifts. This is strong evidence that the expansion of the Universe is accelerating.

One explanation for the acceleration of the Universe's expansion is that there is, after all, a cosmological constant, representing a type of antigravity force that is part of the fabric of space-time. Because the cosmological constant by definition remains constant with time, the Universe would have experienced this acceleration throughout its history.

Another possible explanation is to suppose that totally empty space, the vacuum, contains energy that drives the acceleration. This is an interesting possibility because theoretical physicists have long discussed the idea of a vacuum energy for reasons based on the behavior of subatomic particles. (This is connected with the concept of quantum fluctuations that you read about in the previous section.) Cosmologists label a universal vacuum energy as **quintessence**. Unlike the cosmological constant, the effect of quintessence would not necessarily remain constant over time.

Whatever the correct explanation might be—a cosmological constant antigravity force, a universal vacuum energy, or something else—the observed acceleration of the Universe's expansion is evidence that a type of unknown energy is spread throughout space. Cosmologists refer to it as **dark energy**, an energy that drives the acceleration of the Universe but does not contribute to the formation of starlight or the cosmic microwave background radiation.

You have learned that acceleration and dark energy were first discovered when astronomers found that supernovae a few billion light-years away were slightly fainter than expected based on their redshifts. The acceleration of the expansion has made those supernovae a bit farther away than implied by their redshifts, so they look fainter. Since that discovery, astronomers have continued to find even more distant type Ia supernovae, some as much as 12 billion ly away (**Figure 18-20**). The most distant of those supernovae are not too faint; they are too bright! That observation reveals even more about dark energy, accelerating expansion, and the Universe's history.

Medium-distant supernovae are farther away than their redshifts would indicate without acceleration, whereas very distant supernovae are closer. Those two facts together confirm a theoretical prediction about the behavior of dark energy. When the Universe was young, galaxies were close together; consequently their gravitational pull on each other was stronger than the effect of dark energy, and the expansion decelerated. As space-time

NASA/ESA/STScI/AURA/NSF/A. Riess, GOODS Team

◀ **Figure 18-20** Follow-up observations of extremely distant galaxies located in the GOODS deep fields have detected type Ia supernovae. The upper images here show the supernovae, and the lower images show the galaxies before the supernovae erupted. These supernovae confirm the original discovery that the expansion of the Universe is accelerating and are distant enough to eliminate concerns that the calibration of type Ia supernovae was wrong.

expanded, eventually it moved the galaxies far enough apart that their gravitational pull on each other became weaker than dark energy, and acceleration began.

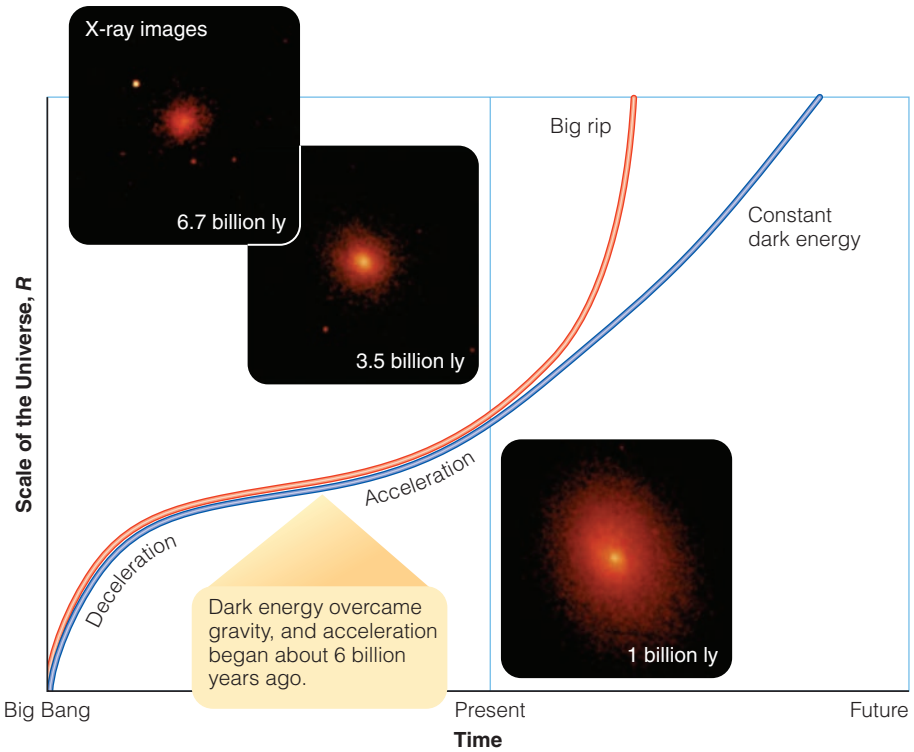
In other words, the observations show that sometime about 8 billion years after the Big Bang (equal to about 60 percent of the current age of the Universe), the Universe shifted gears from deceleration to acceleration, as would be expected when the balance tipped between the opposing effects of gravity and dark energy. The calibration of type Ia supernovae allows astronomers to look back in time and observe the change from universal deceleration to acceleration (**Figure 18-21**).

Step by step, you have been climbing the cosmological pyramid. Each step has been small and logical, but look where it has led you. You now know some of nature's deepest secrets, but you can imagine there are more steps above yet to be found, and more secrets to explore.

18-4b Dark Energy and the Evolution of the Universe

Acceleration solves the problem of the age of the Universe mentioned previously. The Hubble constant H_0 equals 70 km/s/Mpc, and at the beginning of this chapter you calculated the Hubble time, an estimate of the age of the Universe, and found it is about 14 billion years. Later in the chapter you learned that, because gravity would have slowed an originally faster expansion, the current size scale R would have been reached in less time, so the actual age of the Universe should be less than the Hubble time (Figure 18-13). That was a puzzle to astronomers because a few of the globular star clusters are about 13 billion years old, leaving not much room for maneuvering. (As one astronomer said, "You can't be older than your mother.")

The solution lies in the observed acceleration. If the expansion of the Universe has been accelerating, then it must have been



Photos: NASA/CXC/SAO/IoA/S. Allen et al.

◀ **Figure 18-21** X-ray observations of hot gas in galaxy clusters confirm that in its early history, the Universe was decelerating because gravity was stronger than the dark energy. At the point in the past labeled "Acceleration," expansion of the Universe weakened the influence of gravity to the extent that dark energy began to cause acceleration. The data are fit by two model curves of Universe scale versus time. The two curves diverge in the future; the blue curve shows what will happen if the dark energy has a cosmological constant form, whereas the orange curve shows a future of ever-accelerating expansion if the dark energy is quintessence. The data currently are not conclusive but are closer to the cosmological constant model and weigh against quintessence, which means the Universe may not undergo a "big rip." Compare this plot with Figure 18-13 that includes only the effects of gravity.

expanding more slowly in the past, and the average separations of galaxies you can observe now would have taken longer to reach. The latest estimates suggest that acceleration makes the age of the Universe almost 14 billion years, coincidentally about the same as the simple Hubble time estimate (Eq. 18-1 and Figure 18-13). That age is significantly older than the oldest known star clusters, which solves the age problem.

For decades, cosmologists said, "Geometry is destiny." Thinking of open, closed, and flat universe models, they concluded that the density of a model universe determines its geometry, and its geometry determines its fate. By this they meant that if the Universe is open it must therefore expand forever, whereas, if it is closed, it must eventually begin contracting. But that is true only if the behavior of the Universe as a whole is ruled completely by gravity. If dark energy causes an acceleration that can dominate gravity, then geometry is *not* destiny, and depending on the precise properties of dark energy, even a closed universe might expand forever.

The ultimate fate of the Universe depends on the nature of dark energy. If dark energy is described by the cosmological constant, then the force driving acceleration does not change with time, and our flat Universe will expand forever with the galaxies getting farther and farther apart and using up their gas and dust making stars, and stars dying, until each galaxy is isolated, burnt out, dark, and alone. If, however, dark energy is described by quintessence, then the force may be increasing with time, and the Universe might accelerate faster and faster as space pulls the galaxies away from each other, eventually pulling the galaxies apart, then pulling the stars apart, and finally tearing individual atoms apart. This possibility has been called the **big rip**.

Probably there will be no big rip; critically important observations made by the *Chandra X-ray Observatory* have been used to measure the amount of hot gas and dark matter in almost 30 galaxy clusters. Because the distance of each cluster is one of the variables in the calculations, the X-ray astronomers could compare the amount of hot gas and dark matter and solve for distance. The farthest of the clusters is 8 billion ly away. These observations are important for two reasons. First, the redshifts and distance of these galaxies independently confirm the conclusion from the supernova observations that the expansion of the Universe first decelerated but changed to accelerating about 6 billion years ago. Second, the *Chandra* results almost completely rule out quintessence. If dark energy is described by the cosmological constant and not by quintessence, then there will be no big rip (Figure 18-21).

18-4c Inflation: The Enhanced Big Bang Theory

By 1980, accumulating evidence had made the Big Bang theory widely accepted by cosmologists, but it faced several problems that led to the development of a revised Big Bang theory with an important addition.

One of the problems is called the **flatness problem**. The properties of the Universe appear to be close to the dividing line between being open or closed; that is, the geometry of the Universe seems nearly flat, and therefore its density is close to the critical density. If the density of the Universe in the first moments after the Big Bang is calculated based on estimates of its current value, the density back then must have been astoundingly close to the critical value, that is, within a precision of one part in 10^{24} when the Universe was one-billionth of a second old. It is quite remarkable to cosmologists that the density of the Universe would have been so exactly equal to the critical density, and therefore the curvature of the Universe so exactly equal to zero. The flatness problem can be restated as: Why is the Universe so close to perfectly flat when seemingly there is no reason for that?

The second problem with the original Big Bang theory is called the **horizon problem**. This is related to the observed isotropy of the cosmic microwave background radiation. When astronomers correct for the motion of Earth, they see the same intensity and temperature of background radiation in all directions to a precision of better than 1 part in 1000. Yet, when you observe background radiation coming from two points in the sky separated by an angle larger than about 1 degree, you are looking at two bits of matter that seemingly should never have been influenced by each other from the time of the Big Bang up to the time when the radiation was emitted. (The term *horizon* is used in this context because the two spots are said to lie beyond their respective light-travel horizons.) According to the standard Big Bang theory, the material in those two spots of the cosmic background would not have been able to exchange energy and equalize their temperatures. The horizon problem can be restated as: Why did every part of the observable universe have almost exactly the same temperature at the time of recombination, 400,000 years after the Big Bang?

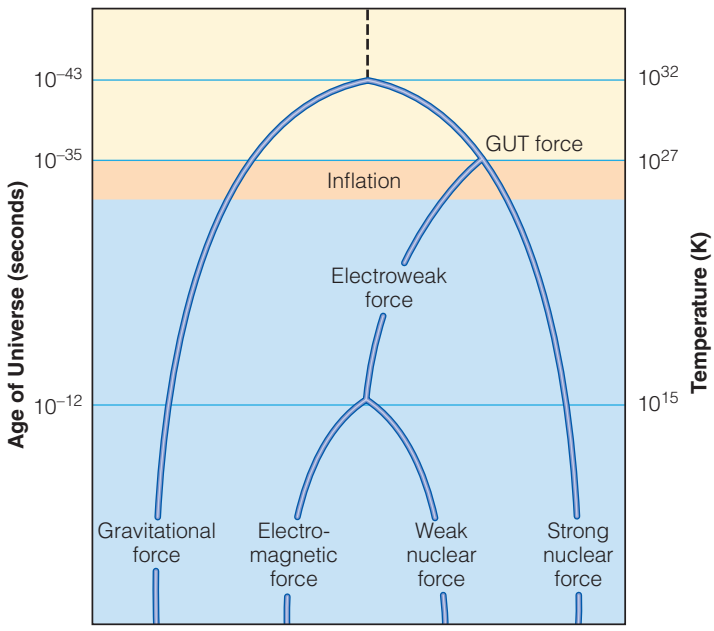
The key to solving these two problems with the original Big Bang theory, as well as other cosmological problems involving subatomic physics, seems to be found in the hypothesis called the **inflationary universe**. That hypothesis predicts there was a sudden and temporary episode of rapid, enormous expansion called *inflation* when the Universe was very young.

To understand the inflationary universe, you need to recall that physicists know of only four forces—gravity, the electromagnetic force, the strong force, and the weak force (Section 8-3a). You are familiar with gravity; you fought it to get out of bed this morning. The electromagnetic force is responsible for making magnets stick to refrigerator doors and cat hair stick to sweaters charged with static electricity, as well as holding electrons in orbit around atomic nuclei and being intimately connected with processes that make light and radiation. The strong nuclear force holds atomic nuclei together, and the weak nuclear force is involved in certain kinds of radioactive decay.

More than a century ago, James Clerk Maxwell showed that the electric force and the magnetic force are intimately connected, and physicists now count them as a single

electromagnetic force. For many years, theorists have tried to unify the other forces; that is, they have tried to describe all the forces of nature as aspects of a single mathematical law. In the 1960s, theorists succeeded in unifying the electromagnetic force and the weak force in what is now called the *electroweak force*. Those two forces can operate effectively as different aspects of a single unified force, but only in very high-energy processes. At lower energies the electromagnetic force and the weak force behave differently. Now, theorists have proposed ways of unifying the electroweak force with the strong force at even higher energies. These new theories are called **grand unified theories (GUTs)**.

According to the inflationary universe hypothesis, the Universe expanded and cooled until about 10^{-36} seconds after the Big Bang, when it became cool enough that the electroweak force and the strong force began to disconnect from each other; that is, they began to behave in different ways (Figure 18-22). Cosmologists calculate that this change would have released tremendous energy, which, during the following 10^{-32} seconds or so, would have inflated the Universe by a size factor of 10^{50} or larger. At the start of inflation, the part of the Universe now observable from Earth is estimated to have been a factor of 10^{35} smaller than a proton, but it suddenly inflated to roughly a meter across and then continued a much slower expansion to its present extent. (Note that, just as with the current expansion of the Universe, nothing actually moved during inflation—rather, space stretched and grew around and between the particles of matter.)



▲ **Figure 18-22** When the Universe was very young and temperatures were extremely high (*top*), the four forces of nature were indistinguishable in behavior. As the Universe began to expand and cool, the forces "separated," meaning that they began to have different characteristics, which released a huge amount of energy and triggered a sudden rapid inflation in the size of the Universe.

In the early 1980s, physicists Alan Guth, Andrei Linde, and others realized that an early episode of rapid expansion can solve both the flatness problem and the horizon problem. The sudden inflation of the Universe would have forced whatever amount of curvature it had before that moment toward a value of zero, just as inflating a balloon makes regions on its surface flatter. Consequently, you now live in a universe that is almost perfectly flat because of that long-ago moment of inflation.

Furthermore, because the entire observable part of the Universe was only about 10^{-58} light-seconds across when the Universe was about 10^{-36} seconds old, that incomprehensibly short interval of 10^{-36} seconds would still have been more than enough time for energy to flow across the region that became the current observable part of the Universe and equalize its temperature before inflation started. As a result, the Universe's background radiation now has almost exactly the same temperature in all directions.

Now you can understand how the existence of dark energy that has caused the Universe's expansion to accelerate in the past few billion years can offer indirect support for the inflation theory. The inflation hypothesis makes a specific prediction that the Universe is very close to exactly flat, so its density must equal the critical density. But baryonic matter plus dark matter make up only about 30 percent of the critical density. Dark energy can make up the difference. As you know, $E = mc^2$ means that energy and matter are equivalent. Thus, the dark energy is equivalent to an amount of mass spread through space. When dark energy is included in the census of matter and energy, the total density of the Universe could equal the critical density, making the Universe flat. But, the proof is in the (raisin) pudding. Is the Universe's geometry flat? Did inflation really occur?

18-4d Fitting It All Together

The inflation-modified Big Bang theory makes several predictions that can be checked. The theory's most basic prediction is that the Universe should be extremely close to exactly flat. Another prediction is that the sudden expansion should have produced gravity waves—in effect, ripples in space-time—which would have left traces in the cosmic microwave background. Astronomers set to work trying to check those predictions.

To verify (or not) the prediction that the Universe is flat, the curvature of space-time over great distances must be determined. One way to measure the curvature of space-time is to compare the angular and linear sizes of things at great look-back times. That type of measurement has yielded impressive results.

As you already know, the background radiation is nearly isotropic; it looks almost exactly the same in all directions after effects of Earth's motion and emission by material in the foreground are taken into account. After those corrections are made to a map of the background radiation, and the average background intensity is subtracted from each spot on the sky, minor irregularities are evident (Figure 18-8). That is, some spots on the sky look a tiny bit hotter and brighter, or cooler and fainter, than the average. Those variations contain lots of information.

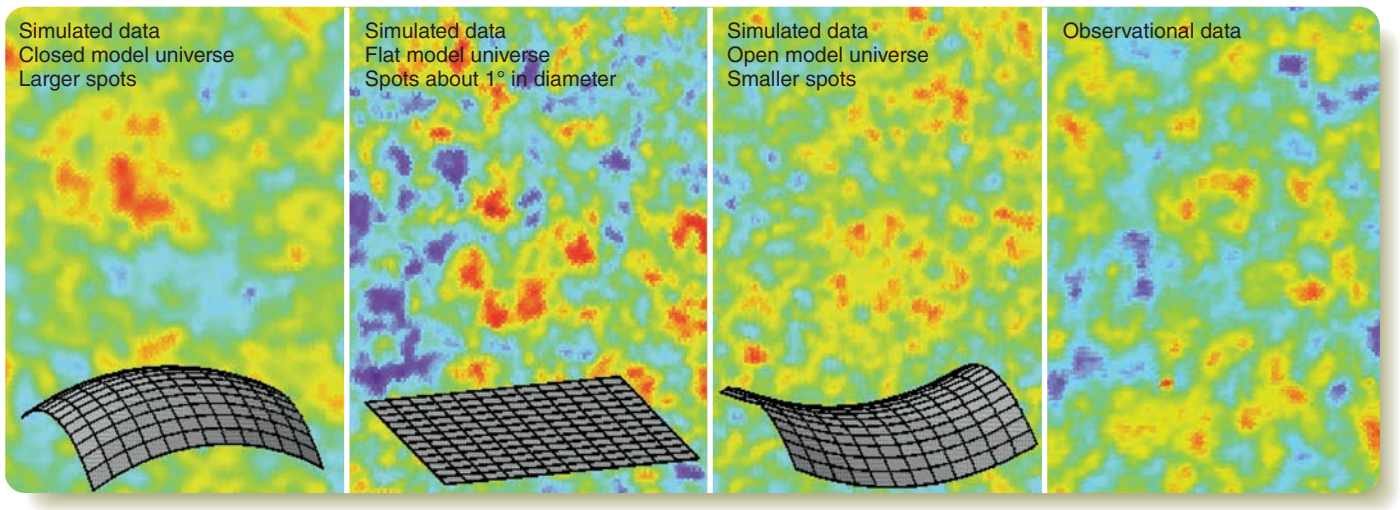
Amazingly enough, the cause of those small variations in the temperature and intensity of the cosmic microwave background radiation was sound waves. It is a **Common Misconception** that explosions in space produce sound. Science-fiction movies imply that sound can travel through a vacuum and that exploding spaceships make big *kabooms.* That's not true now, of course, but the early Universe was dense enough that sound could travel through the gas, so the Big Bang actually made a noise. A theorist described it as "a descending scream, building to a deep rasping roar, and ending in a deafening hiss." The pitch of the sound was about 50 octaves too low for you to hear, but those powerful sound waves did have an effect on the Universe. They produced the irregularities now detectable in the cosmic microwave background radiation. Moreover, the inflation-modified Big Bang theory makes specific predictions about the linear sizes of the variations in the background radiation that can be compared with their observable angular sizes.

Observations made by the *COBE* satellite in 1992 detected the largest-scale variations (Figure 18-8), but the sizes of the smaller variations are critical for testing the theory. Extensive measurements of the spatial distribution of the cosmic background radiation were made using the *Wilkinson Microwave Anisotropy Probe (WMAP)* launched in 2001 and the *Planck* space telescope launched in 2009. Teams of astronomers also sent automated telescopes high in the atmosphere with balloons, while others observed from a facility at the South Pole. An all-sky map of temperature variations in the cosmic background radiation based on 15 months of *Planck* observations is displayed in the image that opens this chapter.

Cosmologists calculate that the most common diameter for the irregularities in the background radiation should be about 1 degree if the Universe is flat. If the Universe is positively curved, the most common irregularities would be larger than that, and if the Universe is negatively curved, they should be smaller. Careful measurements of the size of the variations in the cosmic background radiation show that the observations fit the predictions very well for a flat universe, as you can see in **Figure 18-23**. The fact that the Universe is flat, meaning space-time has no overall curvature and the actual density equals the critical density, indirectly confirms the existence of dark energy, because something other than baryonic matter and dark matter must make up the missing 70 percent of the critical density.

The cosmic microwave background observations and analyses confirm that spots about 1 degree in diameter are the most common, but spots of other sizes occur as well, and it is possible to plot a graph such as **Figure 18-24**, based on analysis of *WMAP* observations, to show how frequently different sizes of irregularities occur.

The data points in Figure 18-24 follow a wiggling line, and the size and positions of those wiggles tell cosmologists a great deal about the Universe. The details of the curve shows that the Universe is flat, accelerating, and will expand forever. The age of the Universe derived from these data is 13.8 billion years. Furthermore, the smaller peaks in the curve reveal that the Universe contains 4.5 percent baryonic (ordinary) matter,



Courtesy of the BOOMERANG Collaboration

▲ **Figure 18-23** You can see the difference yourself. Compare the observations of the irregularities in the background radiation in the far right panel, made using a balloon-borne telescope, with the three simulations starting from the left. The observed size of the irregularities fits best with cosmological models having flat geometry. Detailed mathematical analysis confirms your visual impression: The Universe is flat.

HOW DO WE KNOW? 18-3

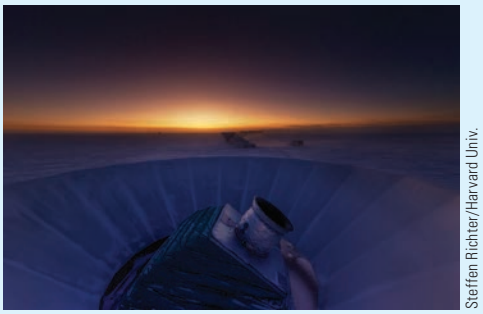
Wishing Doesn't Make It So

The BICEP2 far-infrared telescope shown in the photo was constructed at the South Pole to observe the sky continuously, building an increasingly accurate map of the cosmic microwave background radiation.

In 2014, a team of scientists announced that they had detected a "twist" pattern in the BICEP2 data that would have been produced by gravity waves rushing through the hot gas of the Big Bang. The discovery was exciting because the theory of inflation predicts that gravity waves were present during the Big Bang. Detecting signs of those gravity waves would have confirmed that inflation really did occur during the Big Bang. Scientists around the world were elated.

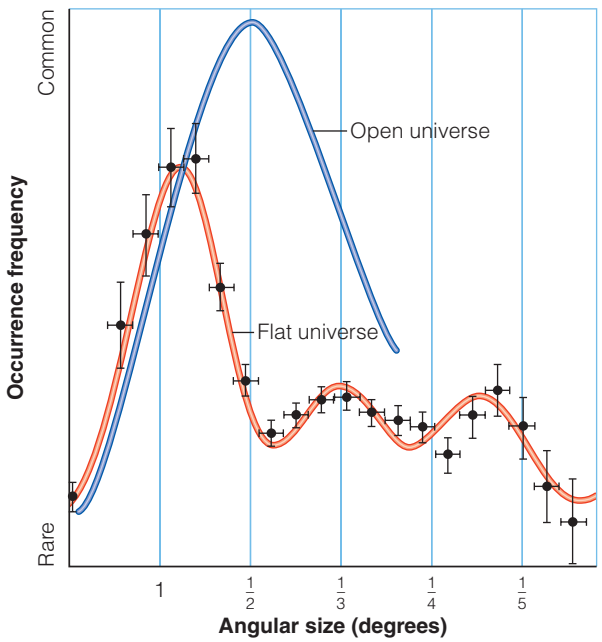
Then, only a few months after that announcement, another team of scientists repeated the analysis and found that the first team may not have corrected properly for the presence of dust in the Milky Way Galaxy. That dust glows at infrared and microwave wavelengths and could partly mimic the pattern that would be produced by gravity waves. Thus, the discovery is inconclusive—the BICEP2 data, when more carefully analyzed, do not show clear evidence of gravity waves, but also do not rule out gravity waves.

Improved detectors, more observations, and more sophisticated analyses should resolve the controversy, but this incomplete story is worth telling for what it shows about how science works. First, new discoveries in science are always checked by other scientists. The more important a result might be, the more quickly and more thoroughly it gets checked. Second, being popular or expected does not make a discovery necessarily true. In science, wishing doesn't make it so.



Steffen Richter/Harvard Univ.

The BICEP2 microwave telescope at the South Pole during twilight, which occurs only twice a year. The Amundsen-Scott South Pole station and the MAPO/Keck Array observatory can be seen in the background.



▲ **Figure 18-24** This graph shows how commonly irregularities of different sizes occur in the cosmic microwave background radiation. Irregularities of about 1 degree in diameter are most common. Models of the Universe that are open or closed are ruled out. The data fit a flat model of the Universe very well. Crosses on data points show the uncertainty in the measurements.

22.7 percent nonbaryonic dark matter, and 72.8 percent dark energy. The Hubble constant is confirmed to be 70 km/s/Mpc. The inflationary theory is confirmed, and the data give more support to the cosmological constant version of dark energy, although quintessence is not quite ruled out. Hot dark matter is ruled out. The dark matter must be cold dark matter to have clumped together rapidly enough after the Big Bang to make the galaxy clusters and superclusters we observe.

Please reread the preceding paragraph. Especially to people who have been working in astronomy for years, that collection of firm facts is mind-blastingly amazing. Studies of the cosmic microwave background radiation and the distribution of galaxies have revolutionized cosmology. At last, astronomers have accurate observations against which to test theories. The values of the basic cosmic parameters are known to a precision of 1 percent or better.

Another prediction of inflation is that gravity waves produced during inflation would have caused a particular "twist" called *polarization* in the microwave background. In 2014, two teams of researchers made conflicting claims about whether this inflation signature was detectable in their respective observations. The rival teams collaborated on a study published in 2015, concluding that interstellar dust emission can explain both data sets and that more sensitive measurements are needed to resolve this issue (**How Do We Know? 18-3**).

On reviewing the observational discoveries and theoretical results obtained in the past two decades, one cosmologist announced, "Cosmology is solved!"; but that might be premature. Cosmologists don't yet know the nature of dark matter or the dark energy that drives the acceleration; so in a sense, more than 95 percent of the Universe is not understood. Nevertheless, although there are further mysteries to be explored, cosmologists are growing more confident that they can describe the origin and evolution of the Universe.

Practicing SCIENCE

How does inflation theory solve the flatness problem?
This question is a perfect example of a hypothesis invented to explain a puzzling piece of evidence.

The flatness problem can be stated as a question: Why is the Universe so flat? After all, the density of matter in the Universe could be anything from zero to infinite, but the observed sizes of variations in the cosmic background radiation show that the Universe is flat, and therefore the average density of the Universe is very close to the critical density. Furthermore, the density must have been astonishingly close to the critical density when the Universe was very young, or it would not be as close as it is now. The inflationary theory solves this problem by proposing that the Universe had a moment of rapid inflation when it was a tiny fraction of a second old. That inflation drove the Universe toward flatness just as blowing up a balloon makes a spot on the balloon flatter and flatter.

Understanding theory in cosmology is critically important, but science ultimately depends on evidence. Now try a related question: ***What is the evidence that the expansion of the Universe is accelerating?***

WHAT ARE WE?

Products

As you climbed the cosmology pyramid, you negotiated many steps. Most were easy, and all were logical. They have carried you up to some sweeping views, and you have traced the origin of the Universe, the formation of the chemical elements, the birth of galaxies, and the births and deaths of stars. You now have a perspective that few humans share.

We humans can count ourselves among the products of two cosmic processes, gravitational contraction and nuclear fusion. Gravity created instabilities in the hot matter created in the Big Bang and triggered the formation of clusters of galaxies. Further contraction of that material formed individual galaxies and then stars within the galaxies. As stars began to shine through the Universe, nuclear fusion in their cores began to cook hydrogen and helium into the heavier atoms from which humans are made.

As you review the history and structure of the Universe, it is wise to recognize the mysteries that remain; but note that they are mysteries that may be solved rather than mysteries that are unknowable. Only a century ago, humanity didn't know there were other galaxies, that the Universe was expanding, or that stars generate energy by nuclear fusion. Human curiosity has solved many of the mysteries of cosmology and will solve more during your lifetime.

Study and Review

SUMMARY

18-1 Introduction to the Universe

- **Cosmology** is the study of the nature, origin, and evolution of the Universe. Astronomers and physicists who do research in cosmology are called **cosmologists**.
- Cosmologists conclude that the Universe cannot have an edge because an edge would introduce logical inconsistencies. In other words, an edge to the Universe does not make sense. If the Universe has no edge, then it cannot have a center.
- If the Universe were infinite in extent, infinite in age, and static, then every spot on the sky would glow as brightly as the surface of a star, and night would be as bright as day. This problem, commonly labeled **Olbers's paradox**, implies that the Universe had a beginning. The **observable universe** is limited by look-back time. It may be only a tiny portion of the entire Universe, which could be infinite.
- Edwin Hubble's 1929 discovery that the redshift of a galaxy is proportional to its distance, now known as the Hubble law, shows that most galaxies appear to be moving away from each other. That phenomenon is called the **expanding universe**.

18-2 The Big Bang Theory

- Tracing the expansion of the Universe backward in time leads to imagining an initial high-density, high-temperature state commonly called the **Big Bang**. A rough estimate of the age of the Universe based on the currently observed expansion rate (Hubble constant) H_0 is called the **Hubble time** τ_H.

$$\tau_H = \frac{10^{12}}{H_0} \text{ years} \qquad \text{(Eq. 18-1)}$$

- The **cosmic microwave background radiation** is blackbody radiation, with a temperature of 2.725 K, spread nearly uniformly over the entire sky. This radiation originated from the Big Bang and has since been redshifted by a factor of 1100.
- During the earliest seconds of the Universe, matter and **antimatter** particles continuously flashed in and out of existence. A slight excess of ordinary matter remained after most of the matter and antimatter particles annihilated each other.
- As the Universe expanded, it cooled. When cool enough, nuclear fusion could then convert some of the hydrogen into helium, heavier atoms were not produced because no stable nuclei exist with atomic weights of 5 or 8.
- After **recombination**, which occurred when the Universe was about 400,000 years old, the Universe expanded in darkness until the first stars came into existence. This period, which model calculations indicate lasted about 400 million years, is called the **dark age**. Astronomers have observed signs of **re-ionization** caused by the ultraviolet light from first generation of stars after the dark age.

18-3 Space and Time, Matter and Gravity

- The fact that the universe is **isotropic** and **homogeneous**. in its large structural features leads to the **cosmological principle**, the idea that the Universe has no special places. Except for minor local differences, every place is the same, and the view from every place is the same.
- Einstein's theory of general relativity explains that cosmic redshifts are caused by wavelengths of photons stretching as they travel through expanding space-time.
- **Closed universe** models are finite in size; space-time curves back on itself. **Open universe** models are infinite in size; space-time is curved but does not curve back on itself. **Flat universe** models, intermediate between closed and open universe models, are infinite in size; space-time is not curved. Modern observations show that this model is probably the correct one.
- Open model universes contain less density than the **critical density**, whereas closed model universes contain more. If the Universe is flat, then its density must equal the critical density.
- The amounts of deuterium and lithium-7 in the Universe show that ordinary baryonic matter cannot make up much more than 4 percent of the critical density. Dark matter must consist of **nonbaryonic matter** and makes up less than 30 percent of the critical density. Dark matter may be hypothetical subatomic particles labeled **WIMPs**.
- The large-scale structure of the Universe consists of galaxy **superclusters** arranged in great walls and **filaments**, outlining enormous **voids**. The observed sizes and masses of galaxy clusters disagree with models that assume rapidly moving **hot dark matter** particles, which have particles that do not clump together easily. Instead, the observed sizes and masses agree with models that assume slowly moving **cold dark matter**, which clumps more readily.

18-4 21st-Century Cosmology

- Cosmologists propose that the acceleration of the expansion of the Universe is caused by **dark energy**. Dark energy may be described by Einstein's **cosmological constant**. On the other hand, the strength of dark energy may change with time; cosmologists call that type of dark energy **quintessence**.
- Some models of quintessence predict an ever-accelerating expansion leading eventually to a **big rip**, which eventually would tear apart all the objects, including the atoms, in the Universe.
- Two problems with the Big Bang theory are the **flatness problem** (a large expansion would force the Universe to become flat) and the **horizon problem** (all parts of the Universe should not be the same temperature). The **inflationary universe** solves these problems by proposing that the Universe briefly expanded dramatically within approximately the first 10^{-32} second after the Big Bang. As explained by **grand unified theories (GUTs),** the energy needed to drive inflation is thought to come from "separation" of the electroweak force and the strong nuclear force as the Universe cooled.
- The sudden inflation of the Universe is hypothesized to have magnified tiny quantum mechanical fluctuations in the density of matter and energy. These small differences in density resulted in dark matter, followed by baryonic matter, drawing together after the Big Bang to produce the present-day large-scale structure of the Universe.
- Observations confirm that the Universe is flat and that it contains 4.5 percent baryonic matter, 22.7 percent dark matter, and 72.8 percent dark energy. Using these values to take into account both initial deceleration and later acceleration, the actual age of the Universe is calculated to be 13.8 billion years.

REVIEW QUESTIONS

1. Is cosmology the study of the Universe, the observable universe, or both? How do you know?
2. Is a cosmologist an astronomer? Is an astronomer a cosmologist? Why do you think so?
3. How does the darkness of the night sky tell you something important about the age and size of the observable universe?
4. Explain the differences among the observable universe expanding, the Universe expanding, and the Universe's expansion accelerating.
5. If you are the size of an ant and located on the crust of the loaf of bread shown in Figure 18-5, is your world homogeneous and isotropic? How do you know?
6. What is expanding in Figure 18-5, the raisins or the bread? Formulate a general conclusion about the size of the raisins versus the distance between the raisin versus the rate of expansion of the bread. Then, rewrite the conclusion substituting astronomical terms for the raisins and bread.
7. If you placed this textbook in intergalactic space, far from any star, at what temperature on the Kelvin scale would it eventually come to equilibrium? Why? Would the answer be the same if you could have performed the same experiment 13 billion years ago?
8. Is the cosmic microwave background radiation in the room with you now? How do you know?
9. Give an example of a particle and its antiparticle. Explain Einstein's formula $E = mc^2$ relative to particles and antiparticles in the photon and particle soup.
10. Which two types of particles are involved, and what do they create, during recombination?
11. During the time of re-ionization, which was dominant: dark energy, matter, or radiation? During the time of the "photon and particle soup"? Today?

12. If you accept the cosmological principle, how can Earth be located at the center of the observable universe?
13. Why can't an open universe have a center? How can a closed universe not have a center?
14. In which type of model universe is space-time infinite in extent and positively curved? List all possibilities.
15. In which type of model universe is space-time infinite in extent and open? List all possibilities.
16. What is the fate of a closed universe? In what case would that not be true?
17. In which model universe does the average density of the Universe equal the critical density?
18. Give examples of baryonic matter and of nonbaryonic matter.
19. What evidence shows that the Universe is expanding? What evidence shows that the Universe began with a Big Bang?
20. Why couldn't atomic nuclei exist when the Universe was less than about 2 minutes old?
21. Why are measurements of the current density of the Universe important?
22. What percentage of matter is ordinary matter? What percentage is dark matter? What makes up the rest of the Universe's density?
23. How does the inflationary universe hypothesis resolve the flatness problem? How does that hypothesis resolve the horizon problem?
24. If the Hubble constant were really 100 km/s/Mpc instead of 70 km/s/Mpc, much of what astronomers understand about the evolution of stars and star clusters would have to be wrong. Explain why. (*Hint:* What would the age of the Universe be?)
25. What is the evidence that the Universe was homogeneous during its first 400,000 years?
26. How is the cosmological constant different from quintessence? How are they similar?
27. If the Universe is negatively curved, and dark energy is described by quintessence, what is the fate of the Universe?
28. What is the difference between hot dark matter and cold dark matter? How does this difference affect cosmology?
29. Of the following redshift values—$z = 1000$, $z = 0.15$, $z = 6$—which corresponds to a distance to a nearby galaxy? To a time when the Universe is so young that not even stars had formed? To a quasar?
30. What evidence can you cite that the Universe's expansion is accelerating?
31. What evidence can you cite that the Universe is flat?
32. **How Do We Know?** Reasoning by analogy often helps make complicated systems or abstract ideas easier to understand. Why do you have to be careful when using analogies?
33. **How Do We Know?** The word *believe* has a meaning in normal conversation that does not really apply to scientific work. Why might cosmologists hesitate to use the word *believe* when they talk about the Big Bang? What should they use instead?
34. **How Do We Know?** In science, wishing something to be true does not make it so. In a scientific context, what does make something true?

ACTIVE INQUIRY QUESTIONS

1. If you reject any model of the Universe that has an edge in space because you can't comprehend such a thing, shouldn't you also reject any model of the Universe that has a beginning? Isn't a beginning like an "edge" in time, or is there a difference?
2. Do you think Copernicus would have accepted the cosmological principle? Why or why not?
3. Since the start of the 20th century, we have collectively climbed many steps of the cosmology pyramid. Do you think we are close to the top, that is, close to knowing the final truth?

PROBLEMS

1. Use the data in Figure 18-4 to plot a velocity–distance diagram, calculate the Hubble constant H_0, and estimate the Hubble time. Plot velocity on the vertical axis and distance on the horizontal axis. Circle the data points on the diagram that you used to determine H_0. Is your result similar to the value of H_0 given in this text? Explain why or why not.
2. Look at the data on velocity, distance, and wavelength shifts for the galaxy spectra shown in Figure 18-4. For Hydra, calculate the velocity (v) as a fraction of c, the speed of light. Do a search via the Internet to see what object can go this speed. Explain what this speed actually represents. (*Note:* The speed of light c is 3.0×10^5 km/s.)
3. Examine Figure 18-4. Calculate the wavelength shifts (that is, the $\Delta\lambda$) of the ionized calcium H line in the spectra of the Virgo cluster and the Hydra cluster. Are these redshifts or blueshifts? How do you know? Form general conclusions about interrelationships among v (velocity), $\Delta\lambda$ (wavelength shift), and d (distance) based on your two data points. (*Notes:* The lab wavelength of Ca II H is $\lambda_0 =$ 393.5 nm; the speed of light c is 3.0×10^5 km/s.) (*Hint:* Although cosmological wavelength shifts are not Doppler shifts, you can use the Doppler formula, Eq. 7-3.)
4. Measure the lengths of the two arrows in the left and the right panels of Figure 18-5. Create a table listing the two before expansion values and the two after expansion values. List in the same table the changes in arrow lengths from before to after (that is, the change in the lengths of the arrows pointing to the two raisins). If you are the raisin from which the arrows point, formulate a general conclusion regarding the rate each raisin is moving toward or away from you and the distance to the raisin, using your data in your answer. (*Hint:* It will be more convenient to use the mm side of your ruler.)
5. If a galaxy is 9.0 Mpc from Earth and recedes at 510 km/s, what is H_0? What is the Hubble time? How old would the Universe be, assuming space-time is flat and the Universe has not been accelerating? How would acceleration change your answer?
6. Find the wavelength of maximum intensity for the radiation emitted by the gas at the time of recombination. What color would you see the gas at the time of recombination? (*Hints:* Use Wien's law, Eq. 7-2, and examine Figure 6-3.)
7. Find the wavelength of maximum intensity of the cosmic microwave background radiation observed today. What band of the electromagnetic spectrum is that in? (*Hints:* Use Wien's law, Eq. 7-2, and examine Figure 6-3.)
8. Pretend that galaxies are spaced evenly, 2.0 Mpc apart, and the average mass of a galaxy is $1.0 \times 10^{11}\ M_\odot$. What is the average density of matter in the Universe? Which model universe does this density value support? (*Notes:* The volume of a sphere is $\frac{4}{3}\pi r^3$, and the mass of the Sun is 2.0×10^{30} kg.)
9. Figure 18-13 is based on an assumed Hubble constant of 70 km/s/Mpc. How would you change the diagram to fit a Hubble constant of 50 km/s/Mpc?
10. Hubble's first estimate of what we now call the Hubble constant was 530 km/s/Mpc. His distances were too small by a factor of about 7 because of a calibration error. If he had not had that calibration problem, what value for H would he have obtained?
11. What was the maximum age of the Universe predicted by Hubble's first estimate of the Hubble constant, given in Problem 10?
12. If the Local Supercluster is 75 Mpc in diameter, how long does light take to travel from one side to the other? Express your answer in millions of years.

SENSE OF PROPORTION

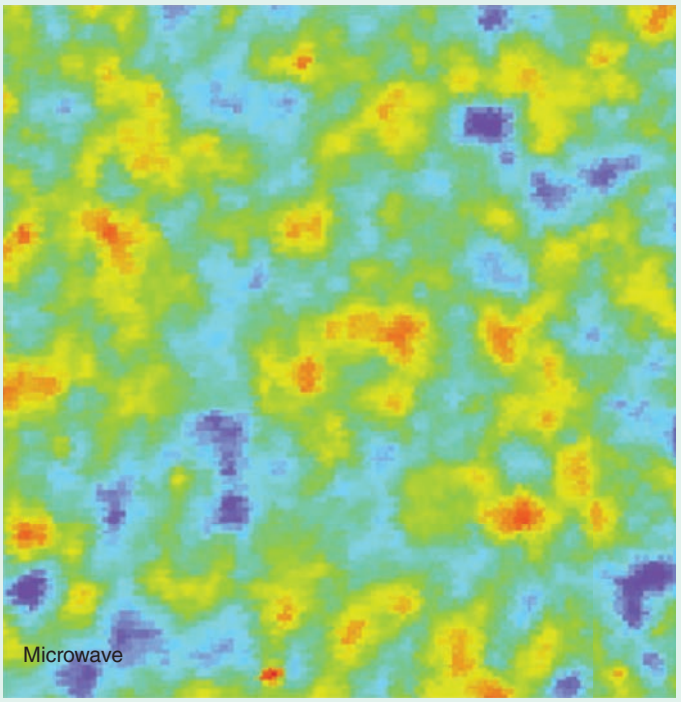
1. If the 13.8 billion-year span from the Big Bang until now were to be represented by a single year of 365 days, how many days before the end of the year would represent the formation of the Earth, Sun, and Solar System, 4.6 billion years before the present? How soon after the start of the year would be recombination, actually 400,000 years after the Big Bang?
2. The distance to the nearest large galaxy, M31, is 2.5 million ly; the distance to the nearest large galaxy cluster, the Virgo cluster is 54 million ly; the distance to the nearest quasar, 3C 273, is 2.4 billion ly. In a model with the distance between the Milky Way and M31 represented by 1 foot, how far away would the Virgo cluster be? 3C 273?

LEARNING TO LOOK

1. Explain why some of the galaxies in this photo have elongated, slightly curved shapes. What do such observations tell you about the Universe?

NASA/ESA/STScI/AURA/NSF/ACS Science Team

2. This image at shows irregularities in the background radiation. Why isn't the background radiation perfectly homogeneous and isotropic? That is, why isn't the background radiation one color? What does the size of these irregularities tell you?



Courtesy of the BOOMERANG Collaboration

3. Examine the enhanced-contrast infrared image in Figure 18-11. Which color do you think represents material most likely to evolve into galaxy clusters and superclusters 1 billion years after the Big Bang? Which color represents material most likely to evolve into intergalactic voids? Which color represents emission from foreground stars, galaxies, and other material?
4. Look at Figure 18-2. Identify the near, far, and very far objects by color, object type, and location in the image. How many objects in the image are stars? (*Hint:* This is a bit of a trick question.)
5. Look at Figure 18-4. Explain what is meant by the speeds listed for each panel in the figure.
6. Look at the *COBE* map of the variations in cosmic microwave background radiation shown in Figure 18-8. Explain the colors of the *COBE* map. Which color represents the locations where stars and galaxies should later form?

19 Astrobiology: Life on Other Worlds

From a painting by Peter Sawyer, Smithsonian Institute

▲ Artist's conception of a scene on the young Earth, more than 3 billion years ago, with stromatolite bacterial mats growing near the shore of an ocean.

This chapter is either unnecessary or vital. If you believe that astronomy is the study of the physical Universe above Earth's clouds, then the previous chapters completed your journey. But if you believe that astronomy is the study not only of the physical Universe but also of your role as a living being in the evolution of the Universe, then everything you have learned so far from this book has been preparation for this final chapter.

As you read this chapter, you will ask four important questions:

- What is life?
- How did life originate on Earth?
- Could life begin on other worlds?
- Can humans on Earth communicate with intelligent species on other worlds?

You won't get more than the beginnings of answers to those questions here, but often in science, asking a good question is more important than getting an immediate answer.

You have explored the Universe from the phases of the Moon to the Big Bang, from the origin of Earth to the death of the Sun. Astronomy is meaningful, not just because it is about the Universe but because it is also about you. Now that you know some astronomy, you can see yourself and your world in a different way. Astronomy has changed you.

Did I solicit thee from darkness to promote me? ADAM, TO GOD, IN JOHN MILTON'S *PARADISE LOST*

AS A LIVING THING, you have been promoted from darkness. The atoms of carbon, oxygen, and other heavy elements that are necessary components of your body did not exist at the beginning of the Universe but were built up by successive generations of stars.

The elements from which you are made are common everywhere in the observable Universe, and planets with Earth-like conditions are almost certainly also common, so it is possible that life began on other worlds. Future explorers may find alien species completely different from any life on Earth. And, it is possible some of those species have evolved to become intelligent. If so, perhaps those other civilizations will be detectable from Earth.

Your goal in this chapter is to try to understand truly intriguing puzzles—the origin and evolution of life on Earth and what that tells you about whether there is life on other worlds. This new hybrid field of study is called **astrobiology**.

19-1 The Nature of Life

What is life? Philosophers have struggled with that question for thousands of years, and it is not possible to answer it completely in one chapter or even one book. An attempt at a general definition of what living things do, distinguishing them from nonliving things, might be: Life is a *process* by which an organism extracts energy from its surroundings, maintains itself, and modifies the surroundings to foster its own survival and reproduction.

One important observation is that all living things on Earth, no matter how apparently different, share certain characteristics in how they perform the process of life.

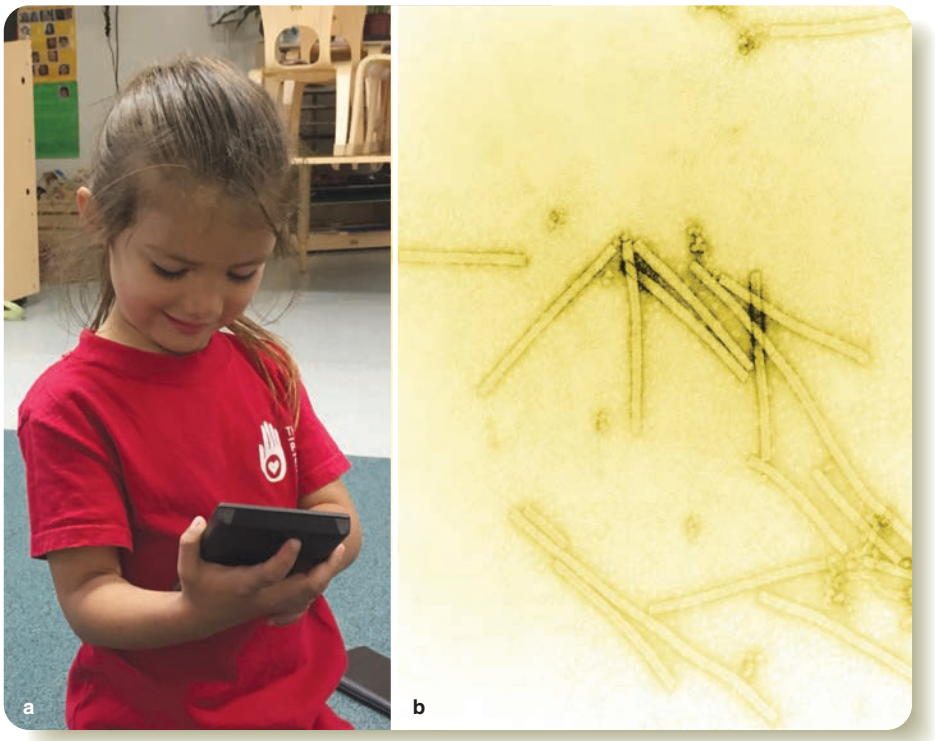
19-1a The Physical Bases of Life

The physical bases of life on Earth are carbon and water (**Figure 19-1**). Because of the way carbon atoms bond to each other and to other atoms, they can join into long, complex, stable chains that are capable, among many other feats, of storing and transmitting information. A large amount of information is necessary to control the activities and maintain the forms of living things. And, in all living things on Earth, the chemical reactions making, breaking, and combining carbon chains take place in liquid water within the cells of living organisms.

Is it possible that life on other worlds could, for example, use silicon atoms in the role of carbon? Silicon is right below carbon in the periodic table (Appendix Table A-14), which means that it shares many of carbon's chemical properties. But life based on silicon rather than carbon seems unlikely to astrobiologists because silicon chains are harder to assemble and disassemble than their carbon counterparts, so they can't be as long and complex nor contain as much information.

Is it possible that the chemistry of life on other worlds could take place in a setting other than water? Some alternatives such as methyl alcohol have been proposed. But carbon compounds dissolve especially easily in water. Also, water has exceptional properties such as a high heat capacity (resistance to temperature change) relative to other cosmically abundant substances that are liquid at the temperatures of planetary surfaces. It's not just because Earth scientists are themselves made of carbon and water that they think carbon and water are crucially important for the existence of life.

Science fiction has proposed even stranger life forms based on, for example, electromagnetic fields and ionized gas, and none of these possibilities can be ruled out. Those hypothetical life-forms make for fascinating speculation, but for now they can't be studied systematically in the way that life on Earth can.



(a) Dana E. Backman; (b) Courtesy of USDA, Beltsville Agricultural Research Center (BARC)

◀ **Figure 19-1** All living things on Earth are based on carbon chemistry. Even the long molecules that carry genetic information, DNA and RNA, have a framework defined by chains of carbon atoms. (a) Sabrina, a complex mammal, contains more than 150 AU of DNA. (b) Each rodlike tobacco mosaic virus contains a single spiral strand of RNA about 0.01 mm long as its genetic material.

This chapter is concerned first with the origin and evolution of life as it is on Earth, based on carbon and water, not because of lack of imagination but because it is the only form of life about which we know anything. From that basis of knowledge we can then sensibly speculate about carbon-based life elsewhere in the Universe. Even carbon- and water-based life has its mysteries. What makes a lump of carbon-based molecules in little bags of water into a living thing? An important part of the answer lies in the transmission of information from one molecule to another.

19-1b Information Storage and Duplication

Most actions performed by living cells are carried out by molecules that are built within the cells. Cells must store recipes for making all those molecules as well as how and when to use them, then somehow pass the recipes on to their offspring.

Study **DNA: The Code of Life** and notice three important points and seven new terms:

1. The chemical recipes of life are stored within each cell as information in *DNA (deoxyribonucleic acid)* molecules that resemble a ladder with rungs that are composed of chemical bases. The recipe information is expressed by the sequence of the rungs, providing instructions to guide specific chemical reactions within the cell.
2. The instructions stored in DNA are genetic information passed along to offspring. DNA instructions normally are expressed by being copied into a messenger molecule called *RNA (ribonucleic acid).* The RNA molecule travels to a location in the cell where its message causes a sequence of molecular units called *amino acids* to be connected into large molecules called *proteins.* Proteins serve as the cell's basic structural molecule or as *enzymes* that control chemical reactions.
3. The DNA molecule reproduces itself when a cell divides so that each new cell contains a copy of the original information. A sequence of DNA that composes one instruction is called a *gene.* Genes are organized into long coiled chains called *chromosomes.* The genes linked on one chromosome are normally passed on to offspring together.

To produce viable offspring, a cell must be able to make copies of its DNA. Surprisingly, it is important for the continued existence of life that the copying process includes mistakes.

19-1c Modifying the Information

Earth's environment changes continuously. To survive, species must change as their food supply, climate, or environmental conditions change. If the information stored in DNA could not change then life would become extinct. The process by which life adjusts itself to changing environments is called **biological evolution**.

When an organism reproduces, its offspring receive a copy of its DNA. Sometimes external effects such as natural radiation alter the DNA during the parent organism's lifetime, and sometimes mistakes are made in the copying process so that occasionally the copy is slightly different from the original. Such a change is called a **mutation**. Most mutations make no difference, but some mutations are fatal, killing the afflicted organisms before they can reproduce. In rare but vitally important cases, a mutation can actually help an organism survive.

These changes in DNA produce variation among the members of a species, and that variation allows the species to adapt to a changing environment. For example, all of the squirrels in the park may look the same, but they carry a range of genetic variation. Some may have slightly longer tails or faster-growing claws. These variations may make almost no difference until the environment changes. For example, if the environment becomes colder, a squirrel with a heavier coat of fur will, on average, survive longer and produce more offspring than its normal contemporaries. Likewise, the offspring that inherit this beneficial variation will also live longer and have more offspring of their own. In contrast, squirrels containing DNA recipes for thin fur coats will gradually decrease in number.

These differing rates of survival and reproduction are examples of **natural selection**. Over time, the beneficial variation increases in frequency, and a species can evolve until the entire population shares the trait. In this way, natural selection adapts species to their changing environments by selecting, from the huge array of random variations, those that would most benefit the survival of the species.

It is a **Common Misconception** that evolution is random, but that is not true. The underlying mechanisms creating variation within each species may be random, but natural selection is not random because progressive changes in a species are directed by changes in the environment.

Practicing SCIENCE

Why is it important that errors occur in copying DNA?

Sometimes the most valuable questions for a scientist to ask are those that challenge what would appear to be common sense.

It seems obvious that mistakes shouldn't be made in copying DNA, but in fact variation is necessary for long-term survival of a species. For example, the DNA in a starfish contains all the information the starfish needs to grow, develop, survive, and reproduce. The information must be passed on to the starfish's offspring for them to survive. That information needs to change, however, if the environment changes. A change in the ocean's temperature may kill the specific shellfish that the starfish eat. If the starfish are unable to digest another kind of food—if all the starfish have exactly the same DNA—they all will die. But if a few starfish have slightly different DNA that gives them the ability to make enzymes that can digest a different kind of shellfish, the species may be able to carry on.

Variations in DNA are caused both by external factors such as natural radiation and by occasional mistakes in the copying process. The survival of life depends on this delicate balance between mostly reliable reproduction and the introduction of small variations in DNA.

Now ponder the opposite question: ***Why does the DNA copying process need to be mostly reliable?***

19-2 Life in the Universe

Life as we know it consists of just the single example of life on Earth. It is OK to think of all life on Earth as being just a single type of life because, as you learned in the previous section, all living things on Earth have the same physical basis: the same chemistry and the same genetic code alphabet. How life began on Earth and then developed and evolved into its present variety is the only solid information you have to work with, when considering what might be possible on other worlds.

Everything currently known about life on Earth indicates that the same natural processes should lead to the origin of life on some fraction of other planets with liquid water. If there is life on other worlds, does it use DNA and RNA to carry the information for life processes, or different molecules playing the same role, or some radically different scheme? There is no way to know unless another example of life is found on another world. If and when that day comes, even if the non-Earthly life is simple one-celled organisms, the discovery will be one of the most important in the history of science. It will complete the journey of human understanding begun in the Copernican revolution of progressive realizations that Earth is not unique.

Courtesy Chip Clark, Australian National Museum of Natural History

▲ **Figure 19-2** This fossil stromatolite from western Australia is more than 3 billion years old, presenting some of the oldest evidence of life on Earth. Stromatolites are formed, layer on layer, by mats of bacteria living in shallow water where they are covered repeatedly by sediments. Look again at the figure that opens this chapter, an artist's conception of a scene on Earth at that time.

19-2a Origin of Life on Earth

It is obvious that the 4.5 billion chemical bases that make up human DNA did not come together in the right order just by chance. The key to understanding the origin of life lies in picturing the processes of evolution running "backward." The complex interplay of environmental factors with the DNA of generation after generation of organisms drove some life-forms to become more sophisticated over time, until they became the unique and specialized creatures alive today. Imagining this process in reverse leads to the idea that life on Earth began with simple forms.

Biologists hypothesize that the first living things would have been carbon-chain molecules able to copy themselves. Of course, this is a scientific hypothesis for which you can seek evidence. What evidence exists regarding the origin of life on Earth?

The oldest fossils are the remains of sea creatures, and this indicates that life began in the sea. Identifying the oldest fossils is not easy, however. Ancient rocks from western Australia that are about 3.5 billion years old contain matlike features that biologists identify as **stromatolites**, fossilized remains of colonies of single-celled organisms that built up layer after layer of trapped sediments (**Figure 19-2**). Fossils this old are difficult to recognize because the earliest living things did not contain easily preserved hard parts like bones or shells and because the individual organisms were microscopic. Thus, the evidence, although scarce and subtle, indicates that simple organisms lived in Earth's oceans less than 1.2 billion years after Earth formed. Stromatolite colonies of microorganisms are more complex than individual cells, so you can imagine there probably were earlier, simpler organisms. How did those first simplest organisms originate?

An important experiment performed by Stanley Miller and Harold Urey in 1952 sought to recreate the presumed conditions in which life on Earth began. The **Miller–Urey experiment** consisted of a sterile, sealed glass container holding water, hydrogen, ammonia, and methane, thought to resemble the young Earth's atmosphere. An electric arc inside the apparatus made sparks to simulate the effects of lightning (**Figure 19-3**).

Miller and Urey let the experiment run for a week and then analyzed the material inside. They found that the interaction between the electric arc and the simulated atmosphere had produced many organic molecules from the raw material of the experiment, including such important building blocks of life as amino acids. (Recall that an organic molecule is simply a molecule with a carbon-chain structure and need not be derived from a living thing: "Organic" does not necessarily imply "biological.")

When the experiment was run again using different energy sources such as hot silica to represent molten lava spilling into the ocean, similar molecules were produced. Even a light source representing the amount of ultraviolet (UV) radiation in sunlight was sufficient to produce complex organic molecules.

Scientists are professionally skeptical about scientific findings (see How Do We Know? 19-2), and they have reevaluated the Miller–Urey experiment in light of new information. According to updated models of the formation of the Solar System and Earth, Earth's early atmosphere probably consisted mostly of carbon dioxide, nitrogen, and water vapor instead of the mix of hydrogen, ammonia, methane, and water vapor assumed by

DNA: The Code of Life

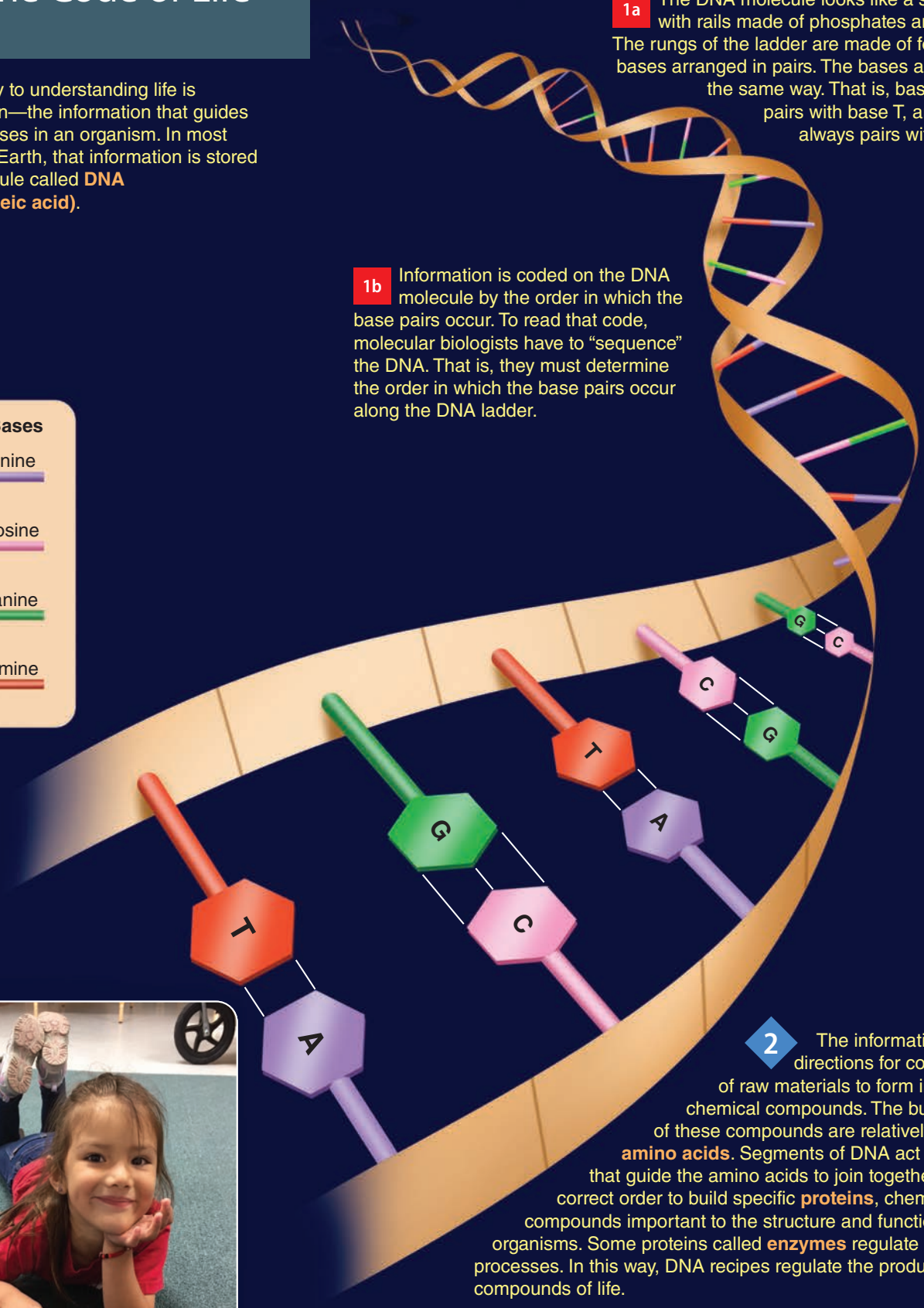
1 The key to understanding life is information—the information that guides all of the processes in an organism. In most living things on Earth, that information is stored in a large molecule called **DNA (deoxyribonucleic acid)**.

1a The DNA molecule looks like a spiral ladder with rails made of phosphates and sugars. The rungs of the ladder are made of four chemical bases arranged in pairs. The bases always pair the same way. That is, base A always pairs with base T, and base G always pairs with base C.

1b Information is coded on the DNA molecule by the order in which the base pairs occur. To read that code, molecular biologists have to "sequence" the DNA. That is, they must determine the order in which the base pairs occur along the DNA ladder.

The Four Bases

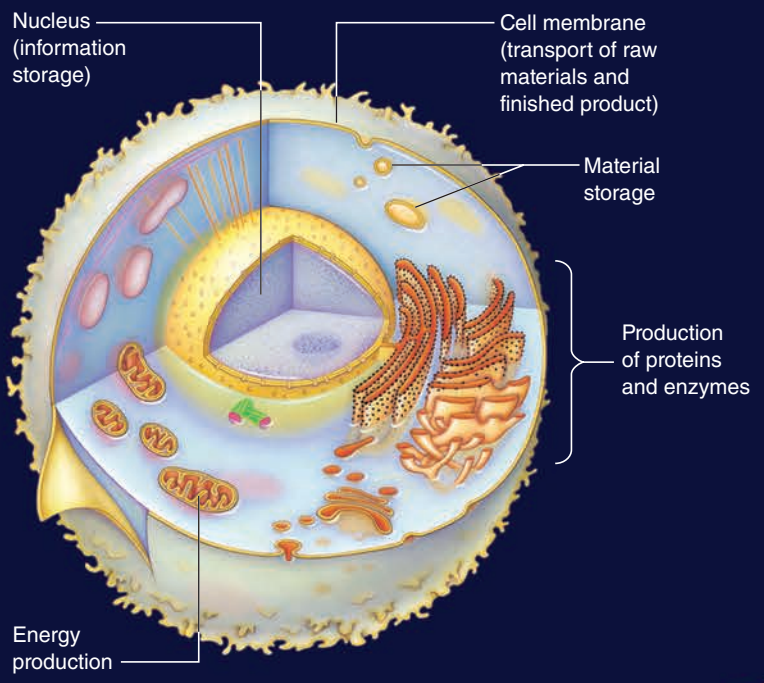
- A — Adenine
- C — Cytosine
- G — Guanine
- T — Thymine



2 The information in DNA is directions for combinations of raw materials to form important chemical compounds. The building blocks of these compounds are relatively simple **amino acids**. Segments of DNA act as templates that guide the amino acids to join together in the correct order to build specific **proteins**, chemical compounds important to the structure and function of organisms. Some proteins called **enzymes** regulate chemical processes. In this way, DNA recipes regulate the production of the compounds of life.

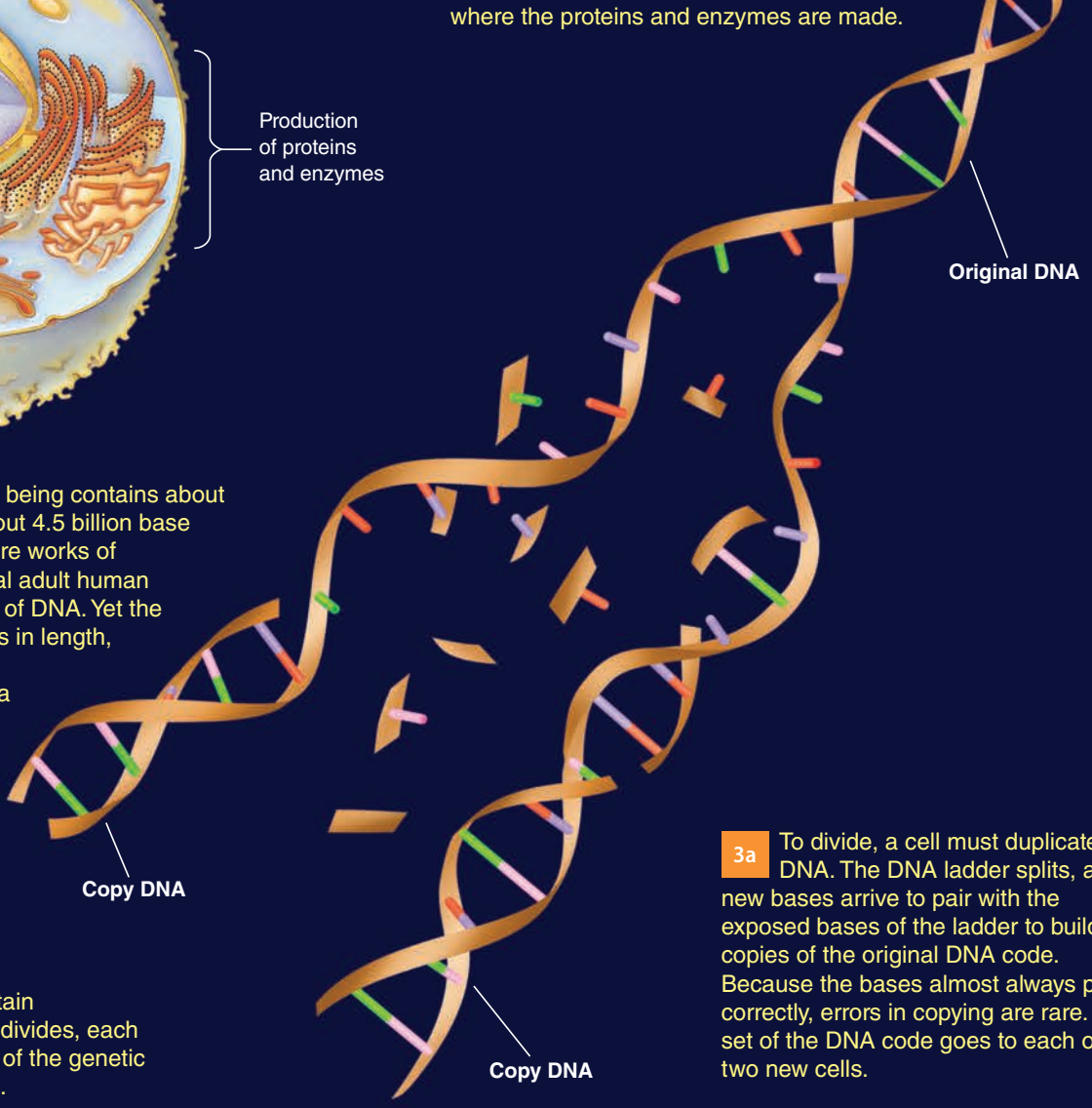
Dana E. Backman

The traits you inherit from your parents, the chemical processes that animate you, and the structure of your body are all encoded in your DNA. When people say, "You have your mother's eyes," they are talking about DNA codes.


Nucleus (information storage)
Cell membrane (transport of raw materials and finished product)
Material storage
Production of proteins and enzymes
Energy production

2a A cell is a tiny factory that uses the information encoded in DNA to produce molecules. Most of the DNA remains safe in the nucleus of a cell, and the code is copied to create a strand of **RNA (ribonucleic acid)**. Like a messenger carrying blueprints, the RNA carries the code out of the nucleus to the work site where the proteins and enzymes are made.

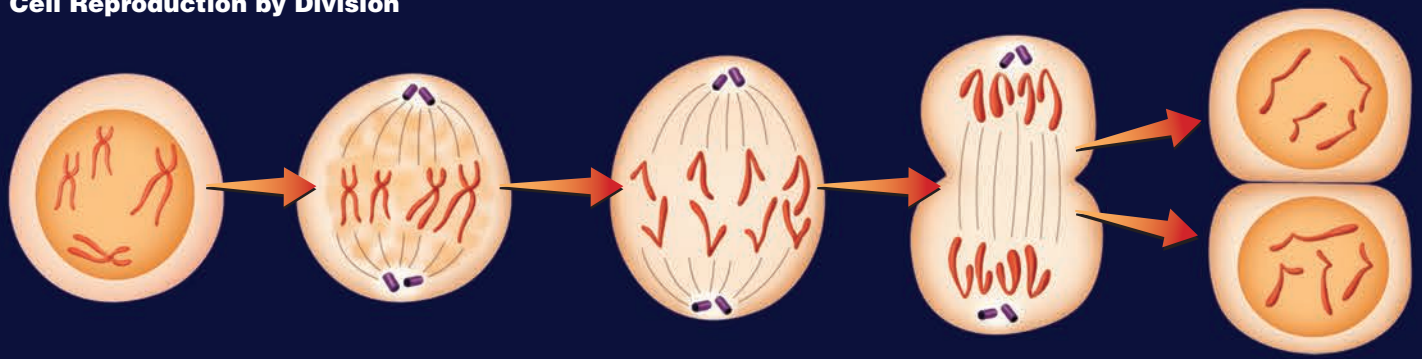
2b A single cell from a human being contains about 1.5 meters of DNA with about 4.5 billion base pairs—enough to record the entire works of Shakespeare 200 times. A typical adult human contains a total of about 600 AU of DNA. Yet the DNA in each cell, only 1.5 meters in length, contains all of the information to create a new human. A clone is a new creature created from the DNA code found in a single cell.



3 DNA, coiled into a tight spiral, makes up the **chromosomes** that contain the genetic material in a cell. A **gene** is a segment of a chromosome that controls a certain product or function. When a cell divides, each of the new cells receives a copy of the genetic information in the chromosomes.

3a To divide, a cell must duplicate its DNA. The DNA ladder splits, and new bases arrive to pair with the exposed bases of the ladder to build two copies of the original DNA code. Because the bases almost always pair correctly, errors in copying are rare. One set of the DNA code goes to each of the two new cells.

Cell Reproduction by Division



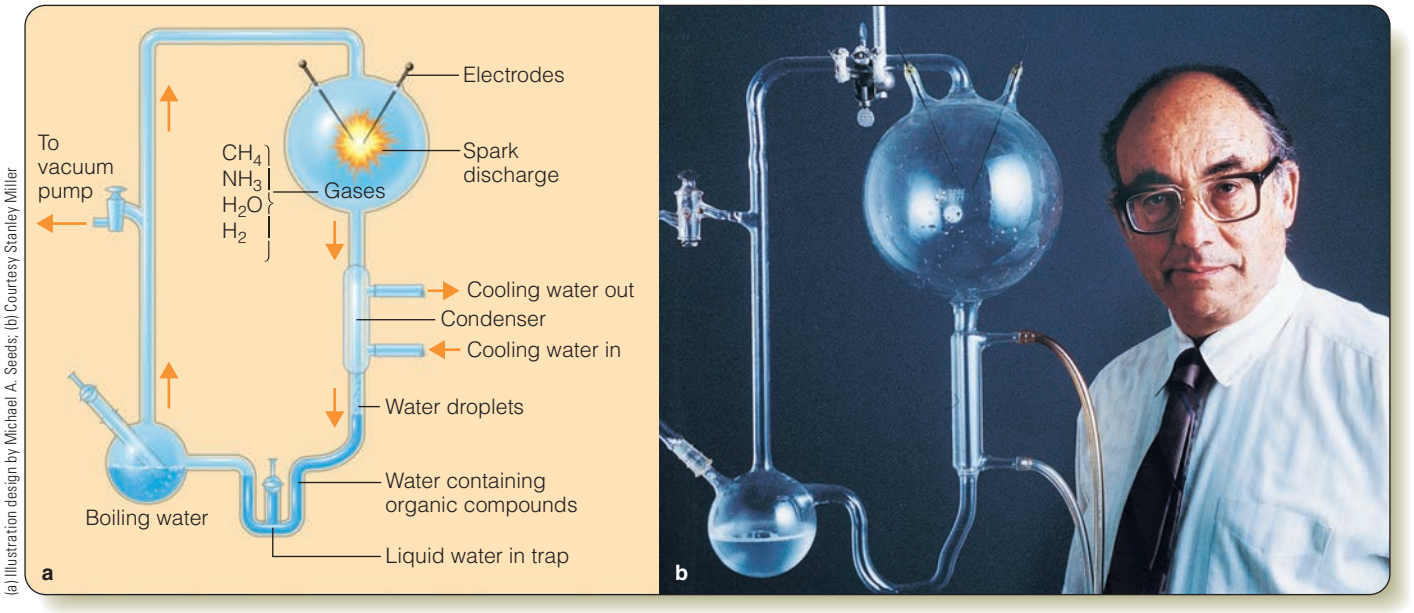
As a cell begins to divide, its DNA duplicates itself.

The duplicated chromosomes move to the middle.

The two sets of chromosomes separate, and . . .

the cell divides to produce . . .

two cells, each containing a full set



(a) Illustration design by Michael A. Seeds; (b) Courtesy Stanley Miller

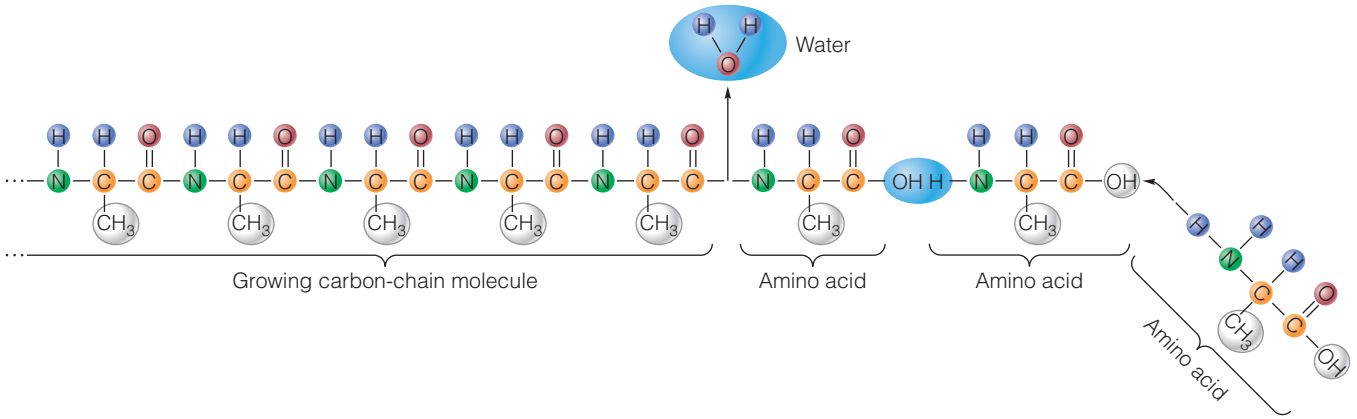
▲ **Figure 19-3** (a) The Miller–Urey experiment circulated gases through water in the presence of an electric arc. This simulation of primitive conditions on Earth produced many complex organic molecules, including amino acids, the building blocks of proteins. (b) Stanley Miller with a Miller–Urey apparatus.

Miller and Urey. When gases corresponding to the newer understanding of the early Earth atmosphere are processed in a Miller–Urey apparatus, lesser, but still significant, amounts of organic molecules are produced.

The Miller–Urey experiment is important because it shows that complex organic molecules form naturally in a wide variety of circumstances. Lightning, sunlight, and hot lava are just some of the energy sources that can naturally rearrange simple common molecules into the complex molecules that make life possible. If you could travel back in time, you would expect to find Earth's early oceans filled with a rich mixture of organic compounds called the **primordial soup**.

Many of these organic compounds would have been able to link up to form larger molecules. Amino acids, for example, link together naturally to form proteins by joining ends and releasing a water molecule (**Figure 19-4**). That reaction, however, does not proceed easily in a water solution. Scientists hypothesize that this step may have been more likely to happen on shorelines or in Sun-warmed tidal pools where organic molecules from the primordial soup could have been concentrated by water evaporation. The production of large organic molecules may have been aided in such semidry environments by being absorbed by clay crystals that could have acted as templates holding the organic subunits close together.

These complex organic molecules were still not living things. Even though some proteins may have contained hundreds of amino acids, they did not reproduce but rather linked and broke apart at random. Because some molecules are more stable than others, and some bond together more easily than others, scientists hypothesize that a process of **chemical evolution**



▲ **Figure 19-4** Amino acids can link together via the release of a water molecule to form long carbon-chain molecules. The amino acid in this hypothetical example is alanine, one of the simplest.

eventually concentrated the various smaller molecules into the most stable larger forms. Eventually, according to the hypothesis, somewhere in the oceans, after sufficient time, a molecule formed that automatically copied itself, as DNA and RNA are able to do under the right circumstances. At that point, the natural selection and chemical evolution of molecules became the biological evolution of living things.

An alternate hypothesis for the origin of life proposes that reproducing molecules may have arrived here from space. Astronomers have found a wide variety of organic molecules in the interstellar medium, and similar compounds have been found inside meteorites (**Figure 19-5**). The Miller–Urey experiment showed how easy it is for complex organic molecules to form naturally from simpler compounds, so it is not surprising to find them in space. Although speculation is fun, the hypothesis that life arrived on Earth from space is presently more difficult to test than the hypothesis that Earth's life originated on Earth.

Whether the first reproducing molecules formed here on Earth or in space, the important thing is that they could have formed by natural processes. Scientists know enough about those processes to feel confident about them, even though some of the steps remain uncertain.

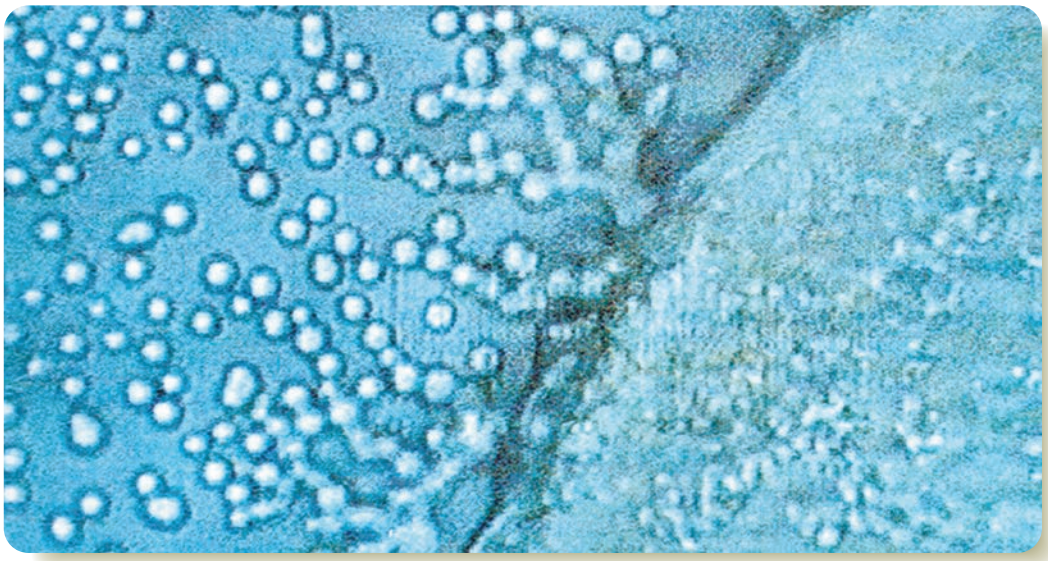
The details of the origin of the first cells are unknown. The structure of cells may have arisen automatically because of the way molecules interact during chemical evolution. If a dry mixture of amino acids is heated, the acids form long, proteinlike molecules that, when poured into water, collect to form microscopic spheres that behave in ways similar to cells (**Figure 19-6**). They have a thin membrane surface, they absorb material from their surroundings, they grow in size, and they divide and bud just as cells do. However, they contain no large molecule that copies itself, so they are not alive. The first reproducing molecule that was surrounded by a protective membrane, resulting in the first cell, would have gained an important survival advantage over other reproducing molecules.

Courtesy Chip Clark, Australian National Museum of Natural History

▲ **Figure 19-5** A piece of the Murchison meteorite, a carbonaceous chondrite that fell near Murchison, Australia, in 1969. Analysis of its interior revealed the presence of amino acids. Whether the first chemical building blocks of life on Earth originated in space is a matter of debate, but the amino acids found in meteorites illustrate how commonly amino acids and other complex organic molecules occur in the Universe, even in the absence of living things.

19-2b Geologic Time and the Evolution of Life

Biologists infer that the first cells must have been simple, single-celled organisms similar to modern bacteria. As you learned previously, evidence of the presence of these kinds of cells are preserved in stromatolites (see Figure 19-2), mineral formations produced by layers of bacteria and shallow ocean sediments.



Sidney Fox and Randall Grubbs

◀ **Figure 19-6** Single amino acids can be assembled into long proteinlike molecules. When such material cools in water, it often forms microspheres, tiny globules with double-layered boundaries similar to cell membranes. Microspheres may have been an intermediate stage in the origin of life between complex but nonliving molecules and living cells holding molecules capable of reproducing genetic information.

Stromatolite fossils are found in rocks with radioactive ages of 3.5 billion years, and living stromatolites still form in some places today.

Stromatolites and other photosynthetic organisms would have begun adding oxygen, a product of photosynthesis, to Earth's early atmosphere. Oxygen tends to disappear from the atmosphere almost as soon as it is released because it readily combines with iron in the soil and ocean water. Geological evidence indicates that Earth's surface iron became saturated with oxygen about 2 to 2.5 billion years ago, after which the proportion of oxygen in the atmosphere began steadily increasing. Oxygen metabolism produces much more energy per mass of food than other reactions, and biologists speculate that this greater efficiency allowed for the development of multicelled organisms at about that same time. Also, an oxygen abundance of only 0.1 percent would have created an ozone screen, protecting organisms from the Sun's UV radiation and later allowing life to colonize the land.

Over the course of eons, the natural processes of evolution gave rise to stunningly complex **multicellular** life-forms with their own widely differing ways of life. It is a **Common Misconception** to imagine that life is too complex to have evolved from such simple beginnings. It is possible because small variations can accumulate, although that accumulation requires great amounts of time.

There is little evidence of anything more than simple organisms on Earth until about 540 million years ago, almost 3 billion years after the earliest signs of life, at which time fossil evidence indicates that life suddenly developing into a wide variety of complex forms such as the trilobites (**Figure 19-7**). This sudden increase in biological complexity is known as the **Cambrian explosion**, at the beginning of what geologists refer to as the Cambrian period.

If you represented the entire history of Earth on a scale diagram, the Cambrian explosion would be near the top of the column, as shown at the left of **Figure 19-8**. The emergence of most animals familiar to you today, including fishes, amphibians, reptiles, birds, and mammals, would be crammed into the topmost part of the chart, above the Cambrian explosion.

If you magnify that portion of the diagram, as shown on the right side of Figure 19-8, you can get a better idea of when these events occurred in the history of life. Humanoid creatures have walked on Earth for about 4 million years. This is a long time by the standard of a human lifetime, but it makes only a narrow red line at the top of the diagram. All of recorded history would be a microscopically thin line at the very top of the column.

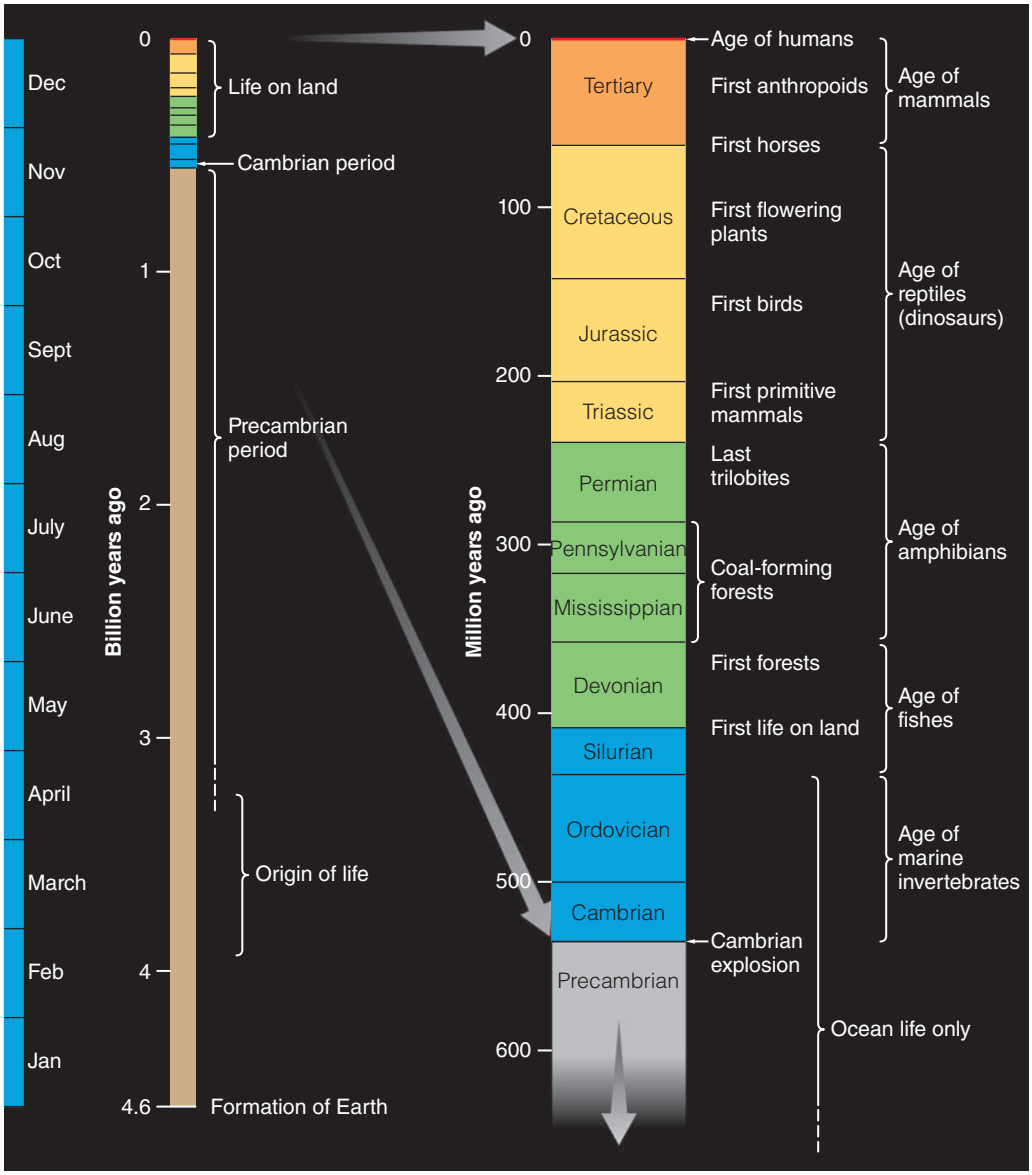
To understand just how thin that line is, imagine that the entire 4.6-billion-year history of the Earth has been compressed onto a yearlong video and that you began watching this video on January 1. You would not see any signs of life until March or early April, and the slow evolution of the first simple forms would take the next six or seven months. Suddenly, in mid-November, you would see the trilobites and other complex organisms of the Cambrian explosion.

You would see no life of any kind on land until the end of November, but once life appeared it would diversify quickly, and by mid-December you would see dinosaurs walking the continents. By the day after Christmas they would be gone, and mammals and birds would be on the rise.

If you watched closely, you might see the first humanoid forms by late afternoon on New Year's Eve, and by late evening you could see humans making the first stone tools. The Stone Age would last until 11:59 pm, after which the first towns, and then cities, would appear. Suddenly things would begin to

(a) Michael A. Seeds; (b) Painting by D. W. Miller, appearing in *American Scientist*, March-April 1997

▲ **Figure 19-7** (a) Trilobites made their first appearance in the Cambrian oceans. The smallest were almost microscopic, and the largest were bigger than dinner plates. This example, about the size of your hand, lived 400 million years ago in an ocean floor that is now a limestone deposit in Pennsylvania. (b) In this artist's conception of a Cambrian sea bottom, Anomalocaris (center-rear and looming at upper right) had specialized organs including eyes, coordinated fins, gripping mandibles, and a powerful, toothed maw. Notice Opabinia at center right with its long snout.



◀ **Figure 19-8** Complex life has developed on Earth only recently. If the entire history of Earth were represented in a time line (*left*), you have to examine the end of the line closely to see details such as life leaving the oceans and dinosaurs appearing. Even on a time line expanded in scale by a factor of about 7 (right column) the age of humans is still only a thin line at the top of the diagram. If the history of Earth were a yearlong video, humans would not appear until the last hours of December 31.

happen at lightning speed. Babylon would flourish, the pyramids would rise, and Troy would fall. The Christian era would begin 14 seconds before the New Year. Rome would fall, and then the Middle Ages and the Renaissance would flicker past. The American and French revolutions would occur one-and-a-half seconds before the end of the video.

By imagining the history of Earth as a yearlong video, you can gain some perspective on the rise of life. Tremendous amounts of time were needed for the first simple living things to evolve in the oceans. As life became more complex, new forms arose more and more quickly as the hardest problems—how to reproduce, how to take energy efficiently from the environment, how to move around—were "solved" by the process of biological evolution. The easier problems, like what to eat, where to live, and how to raise young, were managed in different ways by different organisms, leading to the diversity that is seen today.

Intelligence—that which appears to set humans apart from other animals—may be a unique solution to an evolutionary problem posed to humanity's ancient ancestors. A smart animal is better able to escape predators, outwit its prey, and feed and shelter itself and its offspring. Under certain conditions it seems plausible that evolution might naturally select for intelligence.

19-2c Extremophiles

It is difficult to pin down a range of environments and be sure that life based on carbon and water cannot exist outside those conditions, so long as there is even occasionally some liquid water present. Life has been found in places on Earth previously judged inhospitable, such as the bottoms of ice-covered lakes in Antarctica, far underground inside solid rock, among the cinders at the summits of extinct volcanoes, and in pools

(a) Kris Koenig (Coast Learning Systems); (b) Jim Peaco/National Park Service

▲ **Figure 19-9** Every life form on Earth has evolved to survive in some ecological niche. (a) Wekiu bugs live with the astronomers at an altitude of 13,800 feet atop the Hawaiian volcano Mauna Kea. The bugs inhabit spaces between the icy cinders and eat insects carried up by ocean breezes. (b) Colonies of thermophilic (heat-loving) single-celled organisms thrive around the edge of Yellowstone National Park's Grand Prismatic Spring at temperatures up to 72°C (162°F), producing the green pigments. The blue water in the center of the pool is too hot even for thermophiles.

of scalding-hot water (**Figure 19-9**). An organism that can survive and even thrive in what humans consider an extreme environment is called an **extremophile**. Maybe you have friends like that.

Linguists can figure out the vocabulary of the long-vanished Indo-European language by comparing words in modern languages such as English, Spanish, Russian, Greek, and Hindi that evolved from it. Much the same way, biologists can work out the DNA sequences of ancient species by comparing the sequences of their present-day descendants. This type of analysis indicates that the organism that was ancestor to all life on Earth today most closely resembled present-day simple single-celled organisms known as **archaea**. (The word *archaea* comes from the Greek root *archaios,* meaning ancient.) The most likely common ancestor was an archaea-like extremophile tolerant of high temperatures called a **thermophile** ("heat loving"; look at Figure 19-9b). Some biologists think this is evidence that life began near volcanic vents on the seafloor or in hot rock deep underground. Others suggest that the late heavy bombardment during the end of the Solar System's formation would have repeatedly boiled much of Earth's oceans away. If life had already begun by then, the only organisms that survived this phase of the planet's history to become our ancestors by natural selection would have been heat resistant.

19-2d Life in Our Solar System

Could there be carbon-based life elsewhere in our Solar System? As you learned previously, liquid water seems to be a requirement of carbon-based life, necessary both as the medium for vital chemical reactions and to transport nutrients and wastes. It is not surprising that life developed in Earth's oceans and stayed there for billions of years before it was able to colonize the land. On the other hand, scientists searching for life on other worlds should keep in mind Earth's extremophiles and the harsh conditions in which they thrive.

Many worlds in the Solar System can be eliminated immediately as hosts for water-based life because liquid water is not possible there. The Moon and Mercury are airless, and water would boil away into space immediately. Venus has traces of water vapor in its atmosphere, but it is too hot for liquid water to survive on the surface. The Jovian planets have deep atmospheres, and at a certain altitude it is likely that water condenses into liquid droplets. However, it seems unlikely that life could have originated there. The Jovian planets do not have solid surfaces, so isolated water droplets cannot mingle as they did in Earth's primordial oceans, where organic molecules were able to grow and interact. In addition, powerful downdraft currents in the atmospheres of the giant planets would quickly carry any reproducing molecules that did form there into inhospitably hot lower regions.

Jupiter's moon Europa appears to have a liquid-water ocean below its icy crust, and minerals dissolved in the water could provide a source of raw material for chemical evolution. Europa's ocean is kept warm and liquid by tidal heating. There are evidently also liquid-water zones inside Ganymede, Callisto, and Saturn's moon Enceladus. All of these moons might have been frozen solid at other times in their histories, which would probably have destroyed any living organism that had developed there.

Saturn's moon Titan is rich in organic molecules. Sunlight converts the methane in Titan's atmosphere into organic smog particles that settle to the surface. The chemistry of life

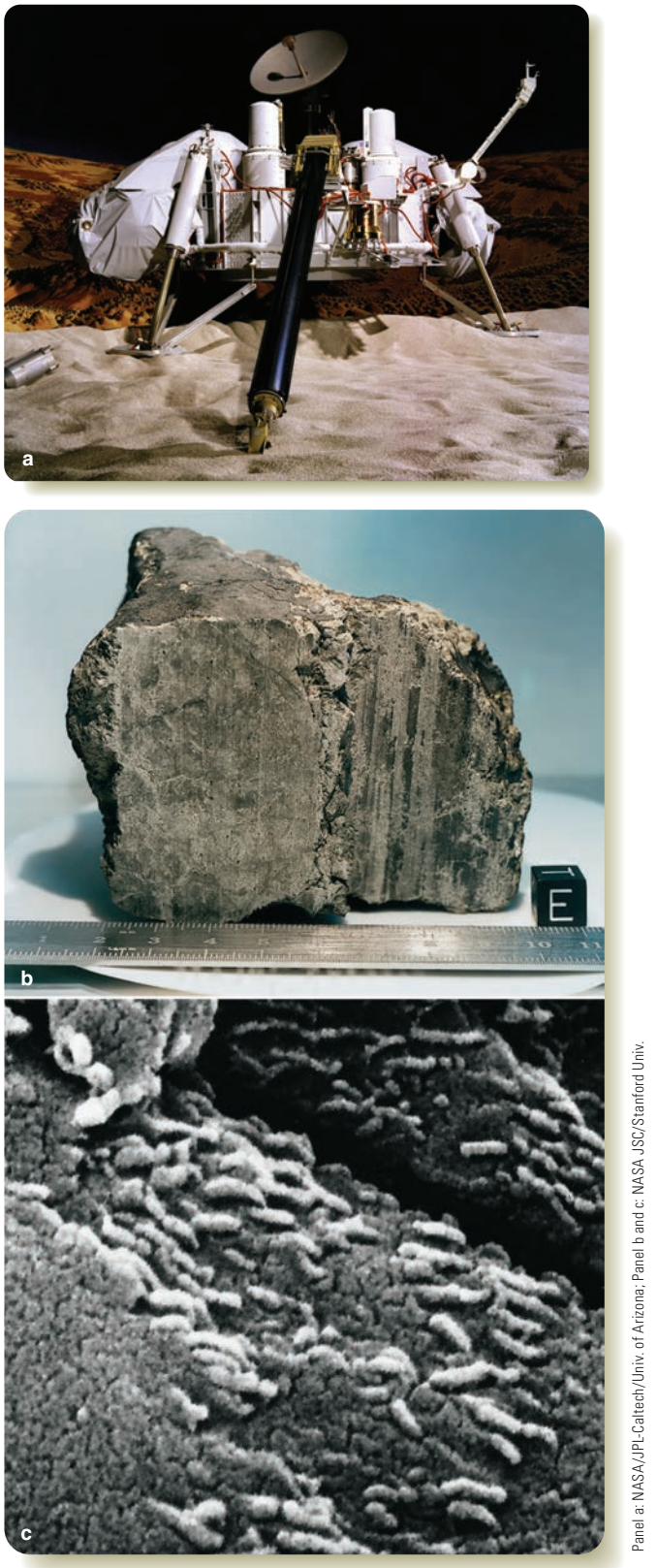
that could have evolved from those molecules and survived in Titan's lakes of methane is unknown. It is fascinating to consider possibilities, but Titan's extremely low temperature of 95 K (−290°F) would make chemical reactions so slow that life processes seem unlikely.

Water containing organic molecules has been observed venting from the south polar region of Saturn's moon Enceladus. It is possible that life could exist in that water under Enceladus's crust, but the moon is small, and its tidal heating might operate only occasionally. Enceladus may not have had plentiful liquid water for the extended time necessary for the rise of life.

Aside from Earth, Mars is the most likely place for life to exist in the Solar System because there is a great deal of evidence that liquid water once flowed on its surface. Even so, results from searches for signs of life on Mars are not encouraging. The robotic spacecraft *Viking 1* and *Viking 2* landed on Mars in 1976 and tested soil samples for living organisms (**Figure 19-10a**). Some of the tests had puzzling semipositive results that scientists now hypothesize were caused by nonbiological chemical reactions in the soil. No evidence clearly indicates the presence of life or even of organic molecules in the Martian soil.

Previously in this chapter you learned that life may have required special circumstances to start on Earth, but once it started, biological evolution allowed life to spread across Earth and adapt to a wide range of conditions. Eventually all niches—even extreme environments—became occupied. Most astrobiologists think this means that, if life begins on a planet, even if the entire environment of the planet later becomes inhospitable, some life could continue to survive. If life still exists on Mars, it may be hidden below ground where there may be liquid water and where UV radiation from the Sun cannot penetrate.

There was a splash of news stories in the 1990s regarding supposed chemical and physical traces of past life on Mars discovered inside a Martian meteorite found in Antarctica (**Figure 19-10b**). Scientists were excited by the announcement, but they employed professional skepticism and immediately began testing the evidence. Their results suggest that the unusual chemical signatures in the rock may have formed by processes that did not involve life. Tiny features in the rock that were originally thought to be fossils of ancient Martian microorganisms could possibly be nonbiological mineral formations instead (**Figure 19-10c**). Although measurements by the *Curiosity* rover have proven conclusively that Mars once had an environment that could have supported Earth-like life, evidence of life on Mars may have to wait until a future rover drills into the soil and discovers signs of metabolizing organisms, or a geologist astronaut scrambles down dry Martian streambeds and cracks rocks open to find fossils.



Panel a: NASA/JPL-Caltech/Univ. of Arizona; Panel b and c: NASA JSC/Stanford Univ.

▲ **Figure 19-10** (a) A *Viking* lander model sits in a simulated Martian environment on Earth. A claw for grabbing samples is at the end of the black arm. (b) Meteorite ALH 84001 is one of a dozen meteorites known to have originated on Mars. (c) Researchers made an unconfirmed claim that ALH 84001 contains traces of life on Mars, including what appear to be fossils of microscopic organisms.

There is presently no compelling evidence for the existence of life in the Solar System other than on Earth. But that means your search can now take you to distant planetary systems.

19-2e Life in Other Planetary Systems

Could life exist in other planetary systems? You already know that there are many different kinds of stars and that many of these stars have planetary systems. As a first step toward answering this question, you can try to identify the kinds of stars that seem most likely to have stable planetary systems where life could evolve.

If a planet is to be a suitable home for living things, it must be in a stable orbit around its sun. That is easy in a planetary system like our own, but planet orbits in binary star systems would be unstable unless the component stars are very close together or very far apart. Astronomers have calculated that, in binary systems with stars separated by intermediate distances of a few AU, the planets should eventually be swallowed up by one of the stars or ejected from the system. Half the stars in the Milky Way Galaxy are members of binary systems, and many of them are therefore unlikely to support life on planets.

Moreover, just because a star is single does not necessarily make it a good candidate for sustaining life. Earth required perhaps as much as 1 billion years to produce the first cells and 4.6 billion years for intelligence to emerge. Massive stars that shine for only a few million years do not meet this criterion. If the history of life on Earth is representative, then stars more massive and luminous than about spectral type F5 last too short a time for complex life to develop. Main-sequence stars of types G and K, and possibly some of the M stars, are the best candidates.

The temperature of a planet is also important, and that depends on the type of star it orbits and its distance from the star. Astronomers have defined a **habitable zone** around a star as a region within which planets that orbit there have temperatures permitting the existence of liquid water. The Sun's habitable zone extends from near the orbit of Venus to the orbit of Mars, with Earth right in the middle. A low-luminosity star has a small and narrow habitable zone, whereas a high-luminosity star has a large and wide one.

Stable planets inside the habitable zones of long-lived stars are the places where life seems most likely, but given the tenacity and resilience of Earth's life forms, there might be other, seemingly inhospitable, places in the Universe where life exists. You should also note that two of the environments considered as possible havens for life—Jupiter's moon Europa and Saturn's moon Enceladus—are in the outer Solar System, far outside the Sun's habitable zone. Those moons have liquid water under their surfaces because of tidal heating due to gravitational interactions with their giant parent planets, a situation that can occur at any distance from a star. Europa and Enceladus show that the conventional definition for habitable zone is probably too limiting.

Practicing SCIENCE

What evidence indicates that life is possible on other worlds? Almost all scientists assume that life exists on other worlds. But that is just an assumption. What is the evidence?

Biologists have imagined, and reproduced in the laboratory, likely physical and chemical processes that, over long time intervals, could have changed simple organic compounds into reproducing molecules inside membranes, the first simple life-forms. Fossil evidence, although meager, indicates that life originated in the oceans at least 3.4 billion years ago, soon after the end of the Solar System's heavy bombardment. That indicates to most astrobiologists that, if conditions are right, life begins on a planet relatively quickly. Finally, evidence indicates that Earth-like planets are common in the Universe.

Now carry the question a little further: ***What conditions do you expect on worlds that host life?***

19-3 Intelligent Life in the Universe

Could intelligent life arise on other worlds? To try to answer this question, you can estimate the chances of any type of life arising on other worlds and then assess the likelihood of that life developing intelligence. If other civilizations exist, it is possible humans eventually may be able to communicate with them. Nature puts restrictions on the pace of such conversations, but the main problem lies in the unknown life expectancy of civilizations.

19-3a Travel Between the Stars

The distances between stars are almost beyond comprehension. The fastest human device ever launched, the *New Horizons* probe, which took 9 years to reach Pluto and now is sailing farther into the Kuiper Belt, will take about 90,000 years to travel 4 light-years, the distance to the nearest star, Proxima Centauri. The obvious way to overcome these huge distances is with tremendously fast spaceships, but even the closest stars are many light-years away.

Nothing can exceed the speed of light, and accelerating a spaceship close to the speed of light takes huge amounts of energy. Even if you travel slower than light, your rocket would still require massive amounts of fuel. If you wanted to pilot a spaceship with a mass of 100 tons (about the size of a fancy yacht) to Proxima Centauri, and you traveled at half the speed of light so as to arrive in 8 years, the trip would require 400 times as much energy as the entire United States consumes in a year. Imagine how much fuel the starship *Enterprise* needs.

These limitations not only make it difficult for humans to leave the Solar System, but they would also make it difficult for aliens to visit Earth. Reputable scientists have studied unidentified flying objects (UFOs) and have never found any evidence that Earth is being visited or has ever been visited by aliens

HOW DO WE KNOW? 19-1

UFOs and Space Aliens

Has Earth been visited by aliens?

Astronomers, planetary scientists, and astrobiologists get asked this question all the time by members of the public. The reason the question is asked so often is that the public has heard that most scientists believe there is life on other worlds. So then, the logic goes, UFOs are probably alien spacecraft, right? Scientists don't make that connection for two reasons.

First, the reputation of UFO sightings and alien encounters does not inspire confidence that those data are reliable. Most people hear of such events in grocery store tabloids, daytime talk shows, or sensational "specials" on viewer-hungry cable networks. You should take note of the low reputation and motivations of the media that report UFOs and space aliens. Most of these reports, like the reports that Elvis is alive and well, are simply made up for the sake of sensation or to make money, and you cannot use them as reliable evidence.

Second, the few UFO sightings that are not made up do not survive careful examination. Most are mistakes and unintentional misinterpretations, committed by honest people, of natural events or human-made objects. It is important to realize that experts have studied these incidents over many decades and found *not even one report* that is convincing to the professional scientific community.

In short, despite false claims to the contrary on TV shows, there is no dependable evidence that Earth has ever been visited by aliens. Remember, that conclusion does not come from a prejudice on the part of scientists against the idea of extraterrestrial life. Most scientists assume there is life on other worlds but are aware that there is no believable evidence for any such life visiting Earth.

In a way, that's too bad. A confirmed visit by intelligent creatures from beyond our Solar System would answer many questions. It would be exciting, enlightening, and, like any real adventure, a bit scary. Most scientists would love to be part of such a discovery. But scientists must professionally pay attention to what is supported by evidence rather than what might be thrilling. There is not yet any direct evidence of even microbial life on other worlds, never mind intelligent extraterrestrial life visiting Earth.

Michael A. Seeds

Flying saucers from space are fun to think about, but there is no evidence that they are real.

(**How Do We Know? 19-1**). Humans are unlikely ever to meet aliens face to face. However, communication by electromagnetic signals across interstellar distances takes relatively little energy.

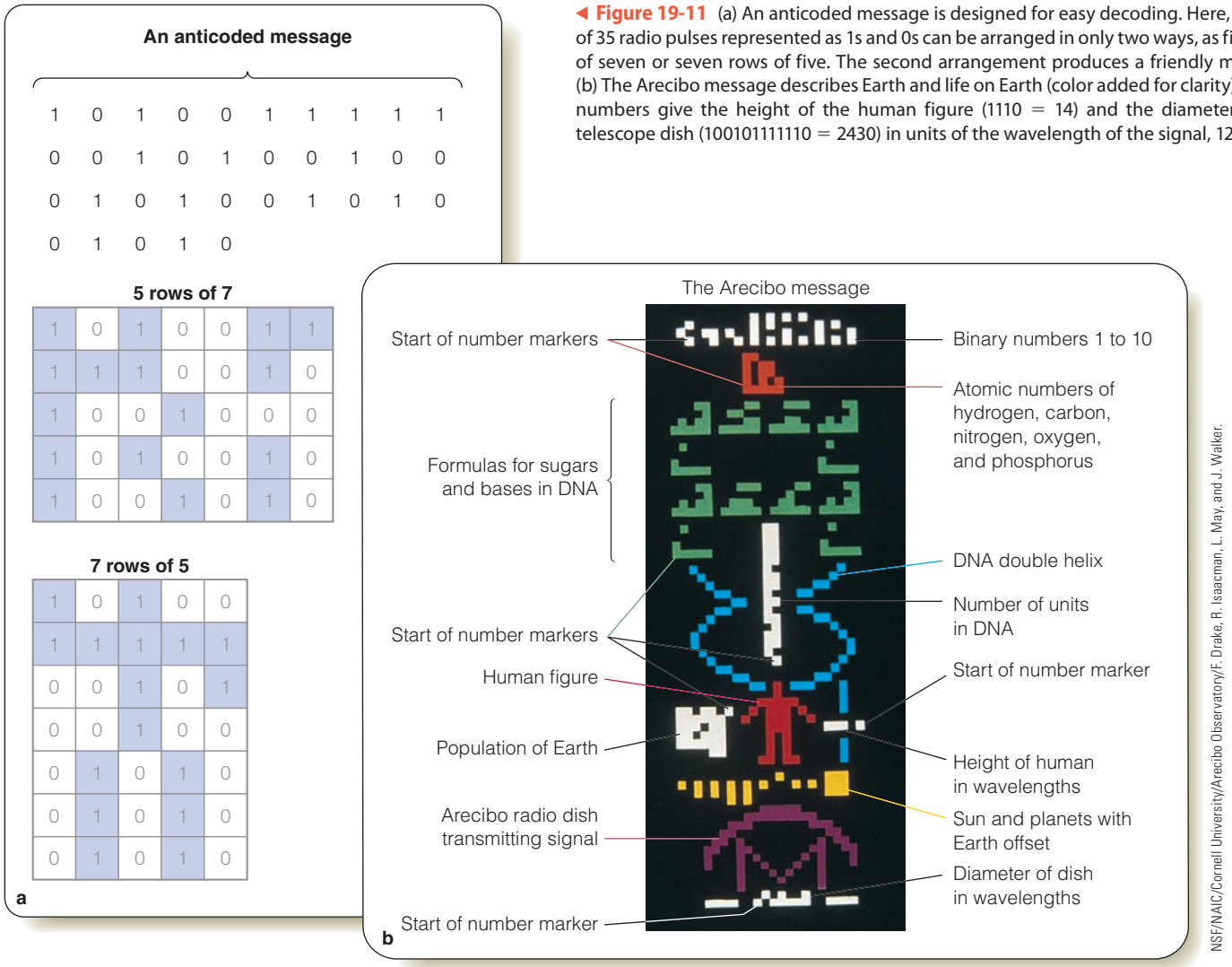
19-3b Radio Communication

Nature puts restrictions on travel through space, and it also restricts the possibility of communicating with distant civilizations by radio. One restriction is based on simple physics: Radio signals are electromagnetic waves and travel at the speed of light. Because of the distances between the stars, the speed of radio waves would severely limit humanity's ability to carry on normal conversations with distant civilizations. Decades could elapse between asking a question and getting an answer.

So rather than try to begin a conversation, one group of astronomers decided in 1974 to broadcast a message of greeting toward the globular cluster M13, 22,000 light-years away, using the Arecibo radio telescope (look back to Figure 6-17). When the signal arrives 22,000 years in the future, alien astronomers may be able to understand it because the message is **anticoded**, meaning that it is designed to be decoded by beings that know nothing about us or our languages. If they are sophisticated enough to build radio telescopes, the hope is that they should be able to decode the transmission. The message is a string of 1679 pulses and gaps. Pulses represent 1s, and gaps represent 0s. The string can be arranged in two dimensions in only two possible ways: as 23 rows of 73 or as 73 rows of 23. The first way produces gibberish, but the second arrangement forms a picture containing information about life on Earth (**Figure 19-11**).

What are the chances that a signal like the Arecibo message would be heard across interstellar distances? Surprisingly, a radio dish the size of the Arecibo telescope, located anywhere in the Milky Way Galaxy, could detect the output from "our" Arecibo. The human race's modest technical capabilities already can put us into cosmic chat rooms.

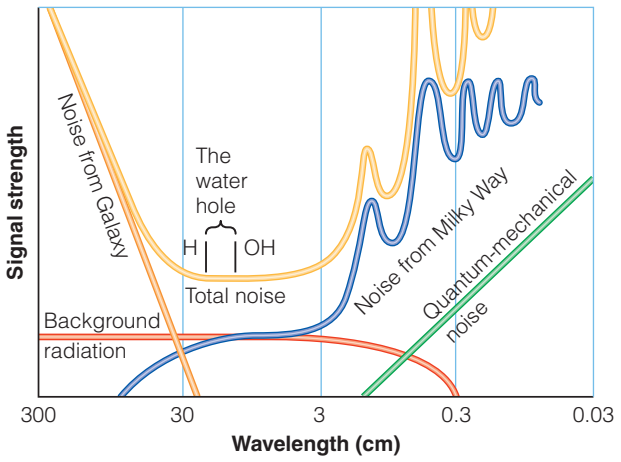
Although the 1974 Arecibo beacon was the only powerful signal sent purposely from Earth to other star systems, Earth is sending out many other signals more or less accidentally. Strong short-wave radio signals, including TV and FM, have been leaking into space for the past 80 years or so. Any civilization within 80 light-years could already have detected Earth's civilization. That works both ways: Alien signals, whether intentional messages of friendship or the blather of their equivalent to daytime TV, could be arriving at Earth now. Groups of astronomers from several countries are pointing radio telescopes at the most likely stars and listening for alien civilizations.


An anticoded message
1 0 1 0 0 1 1 1 1 1
0 0 1 0 1 0 0 1 0 0
0 1 0 1 0 0 1 0 1 0
0 1 0 1 0
5 rows of 7
1 0 1 0 0 1 1
1 1 1 0 0 1 0
1 0 0 1 0 0 0
1 0 1 0 0 1 0
1 0 0 1 0 1 0
7 rows of 5
1 0 1 0 0
1 1 1 1 1
0 0 1 0 1
0 0 1 0 0
0 1 0 1 0
0 1 0 1 0
0 1 0 1 0
a
The Arecibo message
Start of number markers
Formulas for sugars and bases in DNA
Start of number markers
Human figure
Population of Earth
Arecibo radio dish transmitting signal
Start of number marker
Binary numbers 1 to 10
Atomic numbers of hydrogen, carbon, nitrogen, oxygen, and phosphorus
DNA double helix
Number of units in DNA
Start of number marker
Height of human in wavelengths
Sun and planets with Earth offset
Diameter of dish in wavelengths
b
NSF/NAIC/Cornell University/Arecibo Observatory/F. Drake, R. Isaacman, L. May, and J. Walker.

◀ **Figure 19-11** (a) An anticoded message is designed for easy decoding. Here, a string of 35 radio pulses represented as 1s and 0s can be arranged in only two ways, as five rows of seven or seven rows of five. The second arrangement produces a friendly message. (b) The Arecibo message describes Earth and life on Earth (color added for clarity). Binary numbers give the height of the human figure (1110 = 14) and the diameter of the telescope dish (100101111110 = 2430) in units of the wavelength of the signal, 12.6 cm.

Which channels should astronomers monitor? Signals with wavelengths longer than 30 cm would get lost in the background noise of our Milky Way Galaxy, whereas wavelengths shorter than about 1 cm are mostly absorbed in Earth's atmosphere. Between those wavelengths is a radio window that is open for communication. Even this restricted window contains millions of possible radio-frequency bands and is too wide to monitor easily, but astronomers may have thought of a way to narrow the search. Within this broad radio window lie the 21-cm spectral line of neutral hydrogen and the 18-cm line of hydroxyl (OH) (**Figure 19-12**). The interval between those lines has especially low background interference and is named the **water hole** because H plus OH yields water. Any civilizations sophisticated enough to do radio astronomy would know of these lines and might appreciate their significance in the same way as do Earthlings.

A number of searches for extraterrestrial radio signals have been made, and some are now under way. This field of study is known as **Search for Extra-Terrestrial Intelligence (SETI)**, and it has generated heated debate among astronomers, philosophers,



▲ **Figure 19-12** Radio noise from various astronomical sources and Earth's atmospheric opacity make it difficult to detect distant signals at wavelengths longer than 30 cm or shorter than 1 cm. Within the low-noise range, the wavelengths of radio emission lines from H atoms and from OH molecules mark a small interval named the water hole that could be an agreeable channel for interstellar communication.

theologians, and politicians. You might imagine that the discovery of real alien intelligence would cause a huge change in humanity's worldview, akin to Galileo's discovery that the moons of Jupiter do not go around Earth. Congress funded a NASA SETI search for a short time but ended support in the early 1990s. In fact, the annual cost of a major search is only about as much as a single Air Force helicopter, but much of the reluctance to fund searches stems from issues other than cost. Segments of the population, including some members of Congress, considered the idea of extraterrestrial beings to be so outlandish that continued public funding for the search became impossible.

Despite the controversy, the search continues. The NASA SETI project canceled by Congress was renamed Project Phoenix and completed using private funds. The SETI Institute, founded in 1984, managed Project Phoenix plus several other important searches and is currently operating a new radio telescope array in northern California managed by SRI International, partly funded by Paul Allen, one of the cofounders of Microsoft (**Figure 19-13**).

There is even a way for you to help with searches. The Berkeley SETI team (which is separate from the SETI Institute), with the support of the Planetary Society, has recruited about 4 million owners of personal computers that are connected to the Internet. You can download a screen saver that searches data files from the Arecibo radio telescope for signals whenever you are not using the computer. For information, locate the SETI@home project at http://setiathome.berkeley.edu/.

The search continues, but radio astronomers struggle to hear anything against the worsening babble of radio noise from human civilization. Wider and wider sections of the electromagnetic spectrum are being used for Earthly communication, and this, combined with stray electromagnetic radiation from electronic devices including everything from computers to refrigerators, makes hearing faint radio signals difficult. It would be ironic if humans fail to detect faint radio signals from another world because our own world has become too noisy. One alternate search strategy is to look for rapid flashes of laser light at optical or near-infrared wavelengths. Such extraterrestrial signals, if they exist, would have the advantage of being easily distinguished from natural light sources but the disadvantage of being blocked by interstellar dust. Ultimately, the chance of success for any of the searches depends on the number of inhabited worlds in the galaxy.

19-3c How Many Inhabited Worlds?

Given enough time, the searches will find other worlds with civilizations, assuming that there are at least a few out there. If intelligence is common, scientists should find signals relatively soon—within the next few decades—but if intelligence is rare, it may take much longer.

Simple arithmetic can give you an estimate of the number of technological civilizations in the Milky Way Galaxy with which you might communicate, N_C. The formula proposed for discussions about N_C is named the **Drake equation** after the radio astronomer Frank Drake, a pioneer in the search for extraterrestrial intelligence. The version of the Drake equation presented here is modified slightly from its original form:

KEY EQUATION 19-1 The Drake Equation

$$N_C = N_* \cdot f_P \cdot n_{HZ} \cdot f_L \cdot f_I \cdot f_S$$

SETI Institute

◀ **Figure 19-13** Part of the Allen Telescope Array (ATA) near Mount Lassen in California, planned eventually to include 350 radio dishes, each 6 m in diameter, in an arrangement designed to maximize their combined angular resolution. Radio astronomers use the telescope to study galaxies and nebulae of astrophysical interest while SETI researchers employ state-of-the-art computer hardware and software to search for signals from distant civilizations.

The Copernican Principle

Why do astronomers seem so confident that there is life on other worlds?

No message has been received from distant worlds and no life has been detected on any of the worlds visited by probes from Earth. Yet astronomers face this absence of evidence with confidence because of the work of an astronomer who lived 500 years ago.

In Chapter 4, you read about Nicolaus Copernicus and the world he lived in. Astronomers before Copernicus accepted that Earth was the unmoving center of the Universe and that the planets and stars were mysterious lights attached to rotating spheres that carried them around Earth. That made Earth a special place unlike any other place. To better explain these motions, Copernicus proposed that the Earth rotates on its axis and revolves around the Sun. That works much better, but as a consequence it made Earth just another one of the planets. Earth was no longer a special place.

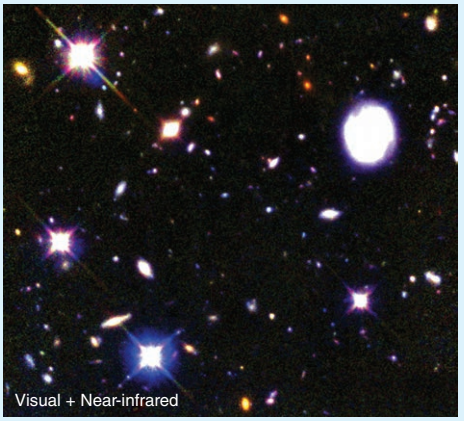
Astronomers have adopted the **Copernican Principle**: Earth is not in a special place. That principle can be extended to assuming that Earth also is not special in other ways.

If Earth is not special, if it is just a planet, then there should be lots of planets like Earth. Of course, many planets are hot, cold, dry, have no atmosphere, or have tremendously deep atmospheres as does Jupiter, but there are a lot of planets. You have learned that a large fraction of stars have planets, that a galaxy contains roughly 100 billion stars, and that there are more than 100 billion galaxies visible with existing telescopes. There must be many planets like Earth.

Although there is no direct evidence of life on other worlds, you have seen in this chapter how life could have originated by natural chemical processes, and how evolution shapes living things to survive and become better adapted. The evidence for that is strong. For astronomers accepting the Copernican Principle, it seems inevitable that life will arise on planets where conditions permit it, and then evolve to become more complex.

You can see the influence of the Copernican Principle in the Drake equation. The factors in the equation follow the logical steps outlined here, and the final factor represents the likelihood that an alien civilization will survive long enough to communicate with other civilizations. Astronomers tend to think that the lack of direct evidence just means we haven't searched long enough or well enough. As Carl Sagan said: "Absence of evidence is not evidence of absence."

When Copernicus said Earth is just one of the planets, he changed the way humanity sees itself, and that change is still rippling through history..



NASA/ESA/STScI/AURA/NSF/HDF-South Team, R. Williams

The *Hubble* Deep Field-South includes many galaxies containing a huge number of stars. There might be more than 1 trillion (10^{12}) planets in this picture.

N_* is the number of stars in our galaxy, and f_P represents the fraction of stars that have planets. If all single stars have planets, f_P is about 0.5. The factor n_{HZ} is the average number of planets in each planetary system suitably located in the habitable zone—meaning, for the sake of the current discussion, the number of planets per planetary system possessing substantial amounts of liquid water. The conventional habitable zone in our system contains Earth's orbit and, arguably, Venus and Mars as well. Europa and Enceladus in our Solar System show that liquid water can exist as a result of tidal heating outside the conventional habitable zone. Thus, n_{HZ} may be larger than had been originally thought. (However, note that, as of this writing, only a handful of Earth-size extrasolar planets have been found orbiting in habitable zones.) The factor f_L is the fraction of suitable planets on which life begins, and f_I is the fraction of those planets where life evolves to intelligence.

Notice that the Drake equation is, in a sense, based on an extension of the idea put forth by Copernicus that Earth is not unique (**How Do We Know? 19-2**). You can put in numbers for the various factors that represent low probabilities or high probabilities, but either way, the implicit assumption of the equation is that Earth, life on Earth, and the human species that desires to make contact with other intelligences in the Universe are single examples of larger sets of such things.

The six factors on the right-hand side of the Drake equation can be roughly estimated, with decreasing certainty as you proceed from left to right. The final factor is extremely uncertain. That factor, f_S, is the fraction of a star's life during which an intelligent species is communicative. If most civilizations last only a short time at a technological level, say for 100 years, there may be none capable of transmitting during the cosmic interval when Earthlings are capable of building radio telescopes to listen for them. On the other hand, a society that stabilizes and remains technologically capable for a long time is much more likely to be detected. For a star with a life span of 10 billion years, f_S might conceivably range from 10^{-8} for extremely short-lived societies to 10^{-3} for societies that survive for ten million years. **Table 19-1**

Table 19-1 The Number of Technological Civilizations per Galaxy

Estimates	Variables	Pessimistic	Optimistic
N_*	Number of stars in a typical large galaxy	2×10^{11}	2×10^{11}
f_P	Fraction of stars with planets	0.1	0.5
n_{HZ}	Number of planets per star that orbit in the habitable zone for longer than 4 billion years	0.01	1
f_L	Fraction of habitable zone planets on which life begins	0.01	1
f_I	Fraction of planets with life on which some species evolves to intelligence	0.01	1
f_S	Fraction of star's existence during which a technological civilization survives	10^{-8}	10^{-3}
N_C	Current number of communicative civilizations per galaxy	2×10^{-4}	1×10^{8}

summarizes what many scientists consider a reasonable range of values for f_S and the other factors.

Example: What would be the estimated number of communicative civilization currently existing in the Milky Way Galaxy, N_C, if values of the component factors are chosen that are between the extremes in Table 19-1? ***Solution:*** If we guess that $f_P = 0.3$, $n_{HZ} = 0.1$, $f_L = 0.1$, $f_I = 0.1$, and $f_S = 10^{-5}$, then:

$$N_C = 2 \times 10^{11} \times 0.3 \times 0.1 \times 0.1 \times 0.1 \times 10^{-5} = 6000$$

When we input values that are neither extremely high nor low, the Drake equation yields an estimate that the Milky Way contains 6000 civilizations able to communicate with each other and us.

If the optimistic estimates are true, there could be a communicative civilization within a few tens of light-years from Earth. On the other hand, if the pessimistic estimates are true, Earth may be the only planet that is capable of communication within thousands of the nearest galaxies.

Practicing SCIENCE

Why does the number of civilizations that could be detected depend on how long civilizations survive at a technological level? Answering this question requires a type of reasoning, used often by scientists, that depends on the timing of events.

Recall from previous chapters that few stars are observed in short-lived life stages because Earthlings' "snapshot" of the Universe catches mostly stars in long-lasting, and therefore more common, parts of their life cycles. Similarly, if you turn a radio telescope to the sky and scan many stars, you would be taking a snapshot of the Universe at a particular time. Broadcasts from other civilizations must be arriving at the time you are observing if you're going to detect them. If most civilizations survive for a long time, there is a much greater chance that you will detect one of them in your snapshot than if civilizations tend to disappear quickly because of, for example, nuclear war or environmental collapse. If most civilizations last only a short time, there may be none transmitting during the cosmically short interval when Earthlings are capable of building radio telescopes to listen for them.

Now consider another aspect of searching for extraterrestrial signals: ***Why might the "water hole" be an especially good frequency band in which to listen?***

WHAT ARE WE?

Matter and Spirit

There are more than 4000 religions around the world, and nearly all hold that humans have a dual nature: We are physical objects made of atoms, but we are also spiritual beings. Science is unable to examine the spiritual side of existence, but it can tell us about our physical nature.

The matter you are made of appeared in the Big Bang and was cooked into a wide range of elements inside stars. Your atoms may have been inside at least two or three generations of stars. Eventually, your atoms became part of a nebula that contracted to form our Sun and the planets of the Solar System.

Your atoms have been part of Earth for the past 4.6 billion years. They have been recycled many times through dinosaurs, stromatolites, fish, bacteria, grass, birds, worms, and other living things. You are using your atoms now, but when you are done with them, they will go back to Earth and be used again and again.

When the Sun swells into a red giant star 6 billion years from now, Earth's atmosphere and oceans will be driven away, and at least the outer few kilometers of Earth's crust will be vaporized and blown outward to become part of the nebula expanding into space away from the white-dwarf remains of the Sun. Your atoms are destined to return to the interstellar medium and become part of future generations of stars and planets.

The message of astronomy is that humans are not just observers: We are participants in the Universe. Among all of the galaxies, stars, planets, planetesimals, and other bits of matter, humans are objects that can think, and that means we can understand what we are.

Is the human race the only thinking species? If so, we bear the sole responsibility to understand and admire the Universe. The detection of signals from another civilization would demonstrate that we are not alone, and such communication would end the self-centered isolation of humanity and stimulate a reevaluation of the meaning of human existence. We may never realize our full potential as humans until we communicate with nonhuman intelligent life.

Study and Review

SUMMARY

19-1 The Nature of Life

- The search for life on other worlds and the investigation of possible habitats for such life is the field of study known as **astrobiology**. Life can be defined as a process by which an organism extracts energy from the surroundings, maintains itself, and modifies the surroundings to promote its survival. The physical bases of life on Earth are carbon and water: The carbon chemistry of Earth life occurs in bags of water (cells).
- Living things have controlling sets of information that can be passed to successive generations. Genetic information for life on Earth is stored in long carbon-chain molecules such as **DNA (deoxyribonucleic acid)**.
- The DNA molecule stores information in the form of chemical bases, which are linked together like the rungs of a ladder. Copied by the **RNA (ribonucleic acid)** molecule, the patterns of bases act as recipes for connecting **amino acid** subunits together to construct **proteins**, including **enzymes**. Proteins are the main structural components, and enzymes control components of the life process.
- The unit of heredity is a **gene**, a piece or several pieces of DNA that in most cases specifies the construction of one particular protein molecule. Genes are connected together in structures called **chromosomes**, which are essentially single, long DNA molecules. When a cell divides, the chromosomes split lengthwise and duplicate themselves so that each of the new cells can receive a copy of the genetic information.
- **Biological evolution** is the process by which species adjusts to changes in the environment. Damage or errors in duplicating the DNA molecule can produce **mutations** that can result in changes to the properties of an organism.
- Genetic variation is essentially random, but **natural selection,** in determining which of these variations are best suited for survival, is controlled by the environment.
- Early fossils called **stromatolites** provide evidence that life began as single-celled organisms in the oceans. The **Miller–Urey experiment** demonstrated the formation of organic molecules under the type of conditions that might have led to the **primordial soup** of Earth's early oceans. **Chemical evolution** concentrated simple molecules into a diversity of larger, stable organic molecules, which were dissolved in the young Earth's oceans. Biological evolution began when molecules developed the ability to make copies of themselves.

19-2 Life in the Universe

- Life did not evolve into more complex, **multicellular** organisms until about 0.5 billion years ago, during a relatively short time known as the **Cambrian explosion**. Life emerged from the oceans only about 0.4 billion years ago. Human intelligence developed over the past 4 million (0.004 billion) years. Anatomically modern humans appeared only about 200,000 years ago.
- Life on Earth requires liquid water and thus a specific range of temperatures and pressures. Organisms that thrive in extreme environments on Earth are called **extremophiles**. Genetic evidence indicates that the common ancestor of all Earth life was a **thermophile**, a heat-tolerant version of present-day single-celled organisms called **archaea**.
- No other celestial object in our Solar System is known to, or appears to, harbor life. Most objects in our Solar System are currently too hot or too cold. Life might have begun on Mars before it became too cold and dry and conceivably could persist in restricted environments. Various moons, such as Europa, Ganymede, and Enceladus, contain liquid water. Titan has abundant organic compounds but no water.
- Main-sequence G and K stars are thought to be the likeliest candidates to host exoplanets with intelligent life because they last long enough for life to potentially form and evolve. The **habitable zone** around a star is the area, within which planets have surface temperatures allowing liquid water on or near their surfaces. However, these may not be the only possible locations for water environments.

19-3 Intelligent Life in the Universe

- Because of the distances involved and energy required, physical travel between stars mean that we are unlikely to ever meet aliens face to face.
- Broadcasting a pulsed radio (or light) beacon would distinguish the signal from naturally occurring emission and would identify the source as a technological civilization. The signal can be **anticoded** in the hope that another intelligent civilization could understand it. One good part of the radio spectrum for communication is called the **water hole**. However, such **Search for Extra-Terrestrial Intelligence (SETI)** programs are looking for radio transmissions from civilizations on other worlds; however, they are hampered by limited computer power and radio noise pollution.
- The number of civilizations in our Milky Way Galaxy that are at a technological level and able to communicate while humans are listening can be estimated by the **Drake equation**. This number is limited primarily by the lifetimes of our and other civilizations. Implicit in the Drake equation is the **Copernican Principle**, that Earth should not be assumed to be a special place.

REVIEW QUESTIONS

1. Explain how astrobiology is a science and not a pseudoscience. (*Hint:* Refer back to Chapter 2 for the definition of *pseudoscience*.)
2. Describe one special quality of water that makes it the basis for life on Earth.
3. Mars and Europa are two Solar System bodies that are considered possible habitats for Earth-like life. Why?
4. How does the DNA molecule produce a copy of itself?
5. What would happen to a life-form if the genetic information handed down to offspring was copied extremely inaccurately?
6. What would happen to a life-form if the genetic information handed down to offspring differs slightly from that in the parents?
7. What would happen to a life-form if the information handed down to offspring was always copied perfectly?
8. Describe an example of natural selection acting on new DNA patterns to select the most advantageous characteristics.
9. Explain how evolution is not random.
10. What evidence do scientists have that life on Earth began in the sea?
11. Define "organic," as in "organic molecule." How is this different from "biological"?
12. Why is liquid water generally considered necessary for the origin of life?
13. Some meteorites contain organic molecules. What does that mean?
14. What is the difference between chemical evolution and biological evolution?
15. What is the significance of the Miller–Urey experiment?
16. Why was Earth's early atmosphere able to support photosynthesis but would not have been able to support animal life?

17. Molecules of which gas were needed in Earth's atmosphere for life to evolve from living in the sea to living on the land? Why?
18. Does intelligence make a creature more likely to survive? Why or why not?
19. Describe one hypothesis for how cells first developed.
20. What is the evidence that the first organisms on Earth most closely resembled today's thermophilic archaea?
21. Name three locations in our Solar System to search for Earth-like life and explain your choices.
22. Why are upper-main-sequence (high-luminosity) host stars unlikely sites for intelligent civilizations?
23. Why is physical travel between stars highly unlikely?
24. How does the stability of technological civilizations affect the probability that humanity can communicate with them?
25. What is the water hole and what role does it have in the search for extraterrestrial civilizations?
26. Why is it hard to make a message that potentially can be understood by alien recipients, and what strategies do scientists use to accomplish this?
27. **How Do We Know?** Why are scientists confident Earth has never been visited by aliens?
28. **How Do We Know?** Why does the Drake equation implicitly assume the Copernican Principle?

ACTIVE LEARNING QUESTIONS

1. Would the atmosphere of an exoplanet be a place where life could begin? Why or why not?
2. If you wanted to find evidence of *intelligent* life on an extrasolar planet, what signatures might you look for in the planet's atmosphere?
3. Do you expect that hypothetical alien recipients of the Arecibo message in Figure 19-11 will be able to decode it? Why or why not?
4. If decades of careful searches for radio signals from extraterrestrial intelligence turn up nothing, what would that mean?

PROBLEMS

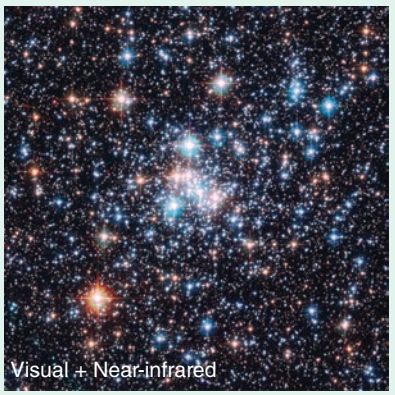
1. A single human cell encloses about 1.5 m of DNA. This length of DNA contains about 4.5 billion base pairs. What is the spacing between these base pairs in units of nanometers? That is, how far apart are the rungs on the DNA ladder? How many base pairs are there per millimeter?
2. If you represent Earth's history by a line that is 1 m long, how long a segment would represent the 400 million years since life first moved onto the land? How long a segment would represent the 4-million-year history of humanoid life?
3. Consider Figure 19-8. What is the ratio of the length of time since the origin of fish to the time since the origin of mammals? What does this value indicate?
4. Suppose a human generation is defined as the average time from birth to childbearing, which is about 20 years long. How many generations have passed in the 200,000 years during which anatomically modern humans have existed?
5. If a star must remain on the main sequence for at least 4 billion years for life to evolve to intelligence, what is the most massive a star that can form and still possibly harbor intelligent life on one of its exoplanets? (*Hints:* Use the formula for stellar life expectancies, Eq. 12-1, and data in Appendix Table A-7.)
6. Mathematician Karl Gauss suggested planting forests and fields in gigantic geometric figures as signals to possible Martians that intelligent life exists on Earth. If Martians had telescopes that could resolve details on Earth no smaller than 1 arc second, how large would the smallest element of Gauss's signal have to be for the element to be visible at Mars's closest approach to Earth? (*Hint:* Use the small-angle formula, Eq. 3-1.) (*Note:* Necessary data are given in Appendix Table A-10.)
7. If you detected radio signals with an average wavelength of 20.000 cm and suspected that they came from a civilization on a distant Earth-like exoplanet, roughly how much of a change in wavelength should you expect to detect as a result of the orbital motion of the distant exoplanet? (*Hint:* Use the Doppler shift formula, Eq. 7-3.) (*Note:* Earth's orbital velocity is 30 km/s.)
8. What is the minimum length of time humans have to wait for a response to the signal sent from Arecibo in 1974 to M13?
9. The first radio broadcast was made on January 13, 1910. It was a live performance at New York City's Metropolitan Opera House. Since that time, how far has that radio broadcast traveled in light-years? In the solar neighborhood there is, on average, one star system per 400 cubic light-years. How many star systems could have heard those opera singers? (*Note:* The volume of a sphere $= \frac{4}{3}\pi r^3$.)
10. Calculate the number of communicative civilizations per galaxy using your own estimates of the factors in Table 19-1.

SENSE OF PROPORTION

1. The DNA in a single cell in your body contains about 4.5 billion base pairs. Imagine building a model of that DNA like a ladder with base pairs for rungs. If you make the rungs 1 inch apart, how long will the ladder be? Express your answer in miles.
2. The Earth is 4.6 billion years old, but human civilization (the building of cities) began only about 10,000 years ago. If you represent the history of Earth with a line 1 mile long (63,360 inches), how long must the line be to represent the history of human civilization?

LEARNING TO LOOK

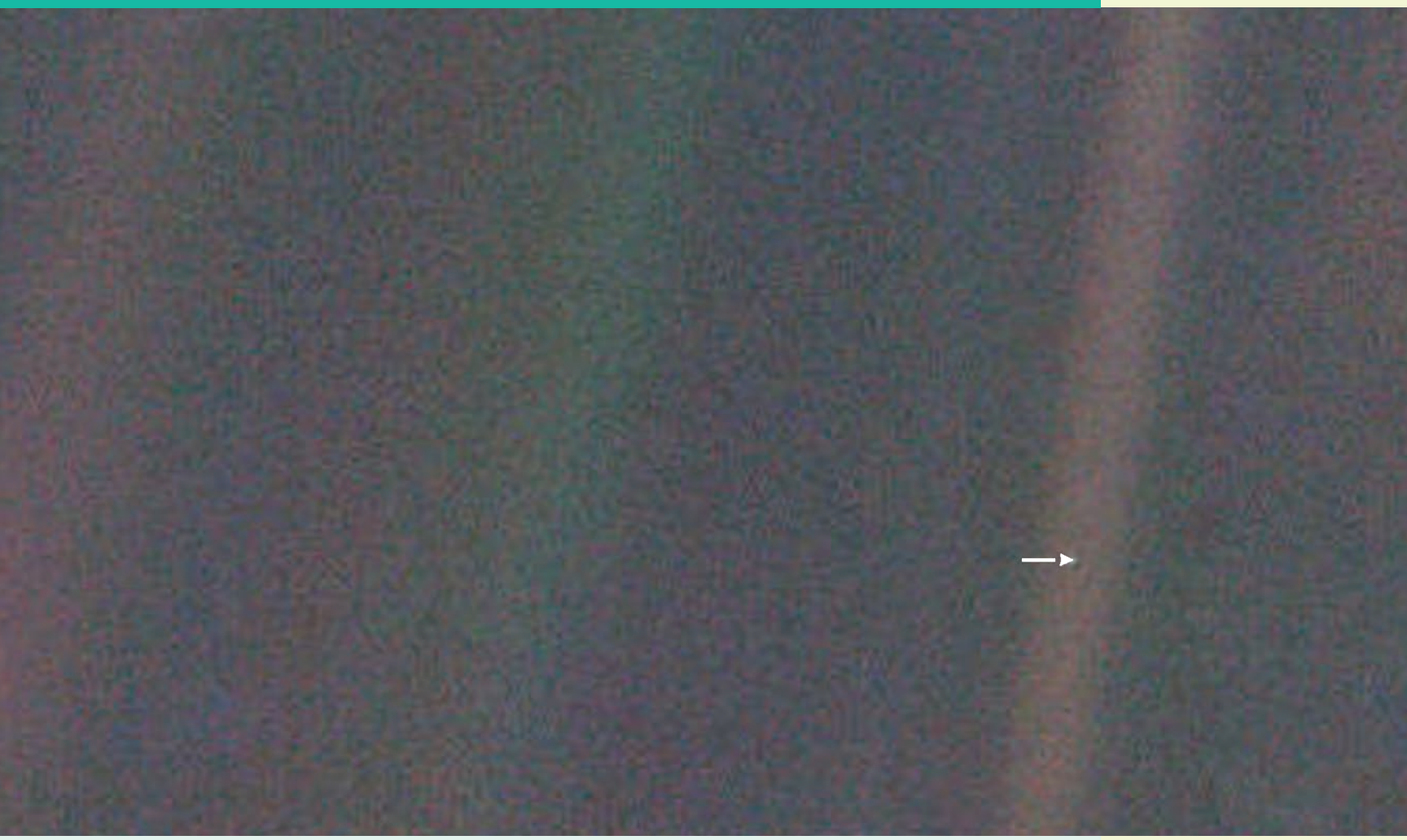
1. Look at Figure 19-11. Since the time we sent the message from Arecibo, what has happened that makes one of the symbols inaccurate?
2. The star cluster shown in this image contains a few red giants as well as main-sequence stars ranging from spectral type B to M. Discuss the likelihood that exoplanets orbiting any of these stars might be home to life. (*Hint:* Estimate the age of the cluster.)
3. If you could search for life in the galaxy shown in this image, would you look among stars in the disk, in the central bulge, in the halo, or in all of those places? Discuss the factors that influence your decision.



NASA/ESA/STScI/AURA/NSF/E. Olszewski (Univ. of Arizona)

ESO

Afterword



NASA

▲ Arrow indicates Earth imaged by *Voyager 1* looking back from the Kuiper Belt, beyond Neptune's orbit. The sunlight beam is an optical effect in the camera.

The aggregate of all our joys and sufferings, thousands of confident religions, ideologies and economic doctrines, every hunter and forager, every hero and coward, every creator and destroyer of civilizations, every king and peasant, every young couple in love, every hopeful child, every mother and father, every inventor and explorer, every teacher of morals, every corrupt politician, every superstar, every supreme leader, every saint and sinner in the history of our species, lived there on a mote of dust, suspended in a sunbeam.

CARL SAGAN (1934–1996)

OUR JOURNEY TOGETHER is over, but before we part company, ponder one final time the primary theme of this book—humanity's place in the physical universe. Astronomy gives us some comprehension of the workings of stars, galaxies, and planets, but its greatest value lies in what it teaches us about ourselves. Now that you have surveyed astronomical knowledge, you can better understand your own position in nature.

To some, the word *nature* conjures up visions of furry rabbits hopping about in a forest glade. To others, nature is the blue-green ocean depths, and still others think of nature as windswept mountaintops. As diverse as these images are, they are all Earth-bound. Having studied astronomy, you can see nature as a beautiful dance of matter and energy, interacting according to simple rules, forming galaxies, stars, planets, mountaintops, ocean depths, forest glades, and people.

Perhaps the most important astronomical lesson is that humanity is a small but important part of the Universe. Most of the Universe is probably lifeless. The vast reaches between the galaxies appear to be empty of all but the thinnest gas, and stars are much too hot to preserve the chemical bonds that seem necessary for life to survive and develop. It seems that only on the surfaces of a few planets, where temperatures are moderate, can atoms link together in special ways to form living matter.

If life is special, then intelligence is precious. The Universe must contain many planets devoid of life, planets where sunlight has shined unfelt for billions of years. There may also exist planets on which life has developed but has not become complex, planets where the wind stirs wide plains of grass and rustles through dark forests. On some planets, creatures resembling Earth's insects, fish, birds, and animals may watch the passing days only dimly aware of their own existence. It is intelligence, human or otherwise, that gives meaning to the landscape.

Science is the process by which Earth's intelligence has tried to understand the physical universe. Science is not the invention of new devices or processes. It does not create home computers, cure the mumps, or manufacture plastic spoons—those are engineering and technology, the adaptation of scientific understanding for practical purposes. Science is the understanding of nature, and astronomy is that understanding on the grandest scale. Astronomy is the science by which the Universe, through its intelligent lumps of matter, tries to understand its own existence.

As the primary intelligent species on this planet, we are the custodians of a priceless gift—a planet filled with living things. This is especially true if life is rare in the Universe. In fact, the rarer life actually is in the Universe, the more overwhelming is our responsibility. We are the only creatures who can take action to preserve the existence of life on Earth; ironically, our own actions are the most serious hazards.

The future of humanity is not secure. We are trapped on a tiny planet with limited resources and a population growing faster than our ability to produce food. We have already driven some creatures to extinction and now threaten others. We are changing the climate of our planet in ways we do not fully understand. Even if we reshape our civilization to preserve our world, the Sun's evolution will eventually destroy Earth.

This may be a sad prospect, but a few factors are comforting. First, everything in the Universe is temporary. Stars die, galaxies die; perhaps the entire Universe will someday end. As part of a much larger whole, we are reminded that our distant future is limited. Only a few million years ago, our ancestors were starting to walk upright and communicate. A billion years ago, our ancestors were microscopic organisms living in the oceans. To suppose that a billion years hence there will be beings resembling today's humans, or that humans will still be the dominant intelligence on Earth, or that human descendants will even exist, is ultimately a conceit.

Our responsibility is not to save our race for all eternity but to behave as dependable custodians of our planet, preserving it, admiring it, and trying to understand it. That calls for drastic changes in our behavior toward other living things and a revolution in our attitude toward our planet's resources. Whether we can change our ways is debatable—humanity is far from perfect in its understanding, abilities, or intentions. However, you must not imagine that we, and our civilization, are less than precious. We have the gift of intelligence, and that is the finest thing this planet has ever produced.

> We shall not cease from exploration
> And the end of all our exploring
> Will be to arrive where we started
> And know the place for the first time.
> —*T. S. Eliot, "Little Gidding"**

*Excerpt from "Little Gidding" in *Four Quartets,* copyright 1942 by T. S. Eliot and renewed 1970 by Esme Valerie Eliot, reprinted by permission of Harcourt Inc. and Faber & Faber Ltd.

Units and Astronomical Data

Introduction

A SYSTEM OF UNITS is based on the three fundamental units for length, mass, and time. By international agreement, there is a preferred subset of metric units known as the *Système International d'Unités* (SI units), commonly called the metric system, which is based on the meter, kilogram, and second. Other quantities, such as density and force, are derived from these fundamental units.

Residents of the United States generally use the (British) imperial system of units (officially used only in the United States, Liberia, and Myanmar but, ironically, not in Great Britain). For example, in imperial units the fundamental unit of length is the foot, composed of 12 inches.

SI units employ the decimal system. For example, a meter is composed of 100 centimeters. Because the metric system is a decimal system, it is easy to express quantities in larger or smaller units as is convenient. You can give distances in centimeters, meters, kilometers, and so on. The prefixes specify the relation of the unit to the meter. Just as a cent is 1/1000 of a dollar, so a centimeter is 1/1000 of a meter. A kilometer is 1000 m, and a kilogram is 1000 g. The meanings of the commonly used prefixes are given in **Table A-1**.

Fundamental and Derived SI Units

THE THREE FUNDAMENTAL SI units define the rest of the units, as given in **Table A-2**.

The SI unit of force is the newton (N), named after Isaac Newton. It is the force needed to accelerate a 1-kg mass by 1 m/s^2, or the force roughly equivalent to the weight of an apple at Earth's surface. The SI unit of energy is the joule (J), the energy produced by a force of 1 N acting through a distance of 1 m. A joule is roughly the energy in the impact of an apple falling off a table.

Exceptions

Units can help you in two ways. They make it possible to make calculations, and they can help you to conceive of certain quantities. For calculations, the metric system is far superior, and it is used for calculations throughout this book.

In SI units, density should be expressed as kilograms per cubic meter, but no human hand can enclose a cubic meter, so that unit does not help you grasp the significance of a given density. This book refers to density in grams per cubic centimeter.

Table A-1 Metric Prefixes

Prefix	Symbol	Factor
giga	G	10^9
mega	M	10^6
kilo	k	10^3
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}

Table A-2 SI (*Système International*) Metric Units

Quantity	SI Unit
Length	meter (m)
Mass	kilogram (kg)
Time	second (s)
Force	newton (N)
Energy	joule (J)
Power	watt (W)

A gram is roughly the mass of a paper clip, and a cubic centimeter is the size of a small sugar cube, so you can easily conceive of a density of 1 g/cm^3, roughly the density of water. This is not a bothersome departure from SI units because you will not have to make complex calculations using density.

For conceptual purposes, this book expresses some quantities in both SI and imperial units. Instead of saying the average adult would weigh 111 N on the Moon, it might be more helpful to some readers for that weight to be expressed as 25 lb. In such cases, the imperial form is given in parentheses after the SI form. For example, the radius of the Moon is 1738 km (1080 mi).

Conversions

To convert from one metric unit to another (from meters to kilometers, for example), you need only look at the prefix. However, converting from metric to English or English to metric is more complicated. The conversion factors are given in **Table A-3**.

Example: The radius of the Moon is 1738 km. What is this in miles? Table A-3 indicates that 1.000 mile equals 1.609 km, so

$$1738 \text{ km} \times \frac{1.000 \text{ mi}}{1.609 \text{ km}} = 1080 \text{ mi}$$

Temperature Scales

In astronomy, as in most other sciences, temperatures are expressed on the Kelvin scale, although the centigrade (or Celsius) scale is also used. The Fahrenheit scale commonly used in the United States is not used in scientific work.

The centigrade scale refers temperatures to the freezing point of water (0°C) and to the boiling point of water (100°C). One degree Centigrade is $^1/_{100}$, the temperature difference between the freezing and boiling points of water, thus the prefix *centi*. The centigrade scale is also called the Celsius scale after its inventor, the Swedish astronomer Anders Celsius (1701–1744).

Temperatures on the Kelvin scale are measured in Celsius degrees from absolute zero (−273.15°C), the temperature of an object that contains no extractable heat. In practice, no object can be as cold as absolute zero, although laboratory apparatuses have reached temperatures lower than 10^{-6} K. The Kelvin scale is named after the Scottish mathematical physicist William Thomson, Lord Kelvin (1824–1907).

The Fahrenheit scale fixes the freezing point of water at 32°F and the boiling point at 212°F. Named after the German physicist Gabriel Daniel Fahrenheit (1686–1736), who made the first successful mercury thermometer in 1720, the Fahrenheit scale is used routinely only in the United States.

It is easy to convert temperatures from one scale to another using the information given in **Table A-4**.

Powers of 10 Notation

Powers of 10 make writing very large numbers much simpler. For example, the nearest star is about 43,000,000,000,000 km from the Sun. Writing this number as 4.3×10^{13} km is much easier.

Very small numbers can also be written with powers of 10. For example, the wavelength of visible light is about 0.0000005 m. In powers of 10, this becomes 5×10^{-7} m.

The powers of 10 used in this notation appear below. The exponent tells you how to move the decimal point. If the exponent is positive, move the decimal point to the right. If the exponent is negative, move the decimal point to the left. For example, 2.0000×10^3 equals 2000, and 2×10^{-3} equals 0.002.

$10^5 = 100{,}000$
$10^4 = 10{,}000$
$10^3 = 1000$
$10^2 = 100$
$10^1 = 10$
$10^0 = 1$
$10^{-1} = 0.1$
$10^{-2} = 0.01$
$10^{-3} = 0.001$
$10^{-4} = 0.0001$

Table A-3 Conversion Factors Between British and Metric Units

1 inch = 2.54 centimeters	1 centimeter = 0.3937 inch
1 foot = 0.3048 meter	1 meter = 39.37 inches = 3.281 feet
1 mile = 1.609 kilometers	1 kilometer = 0.6214 mile
1 slug = 14.59 kilograms	1 kilogram = 0.06852 slug
1 pound = 4.448 newtons	1 newton = 0.2248 pound
1 foot-pound = 1.356 joules	1 joule = 0.7375 foot-pound
1 horsepower = 745.7 joules/s	1 joule/s = 0.001341 horsepower
	1 joule/s = 1 watt

Table A-4 Temperature Scales and Conversion Formulas

	Kelvin (K)	Centigrade (°C)	Fahrenheit (°F)
Absolute zero	0 K	−273°C	−460°F
Freezing point of water	273 K	0°C	32°F
Boiling point of water	373 K	100°C	212°F
Conversions:			
$K = °C + 273$			
$°C = \frac{9}{5}(°F - 32)$			
$°F = \frac{9}{5}(°C) + 32$			

If you use scientific notation in calculations, be sure you correctly enter the numbers into your calculator. Not all calculators accept scientific notation, but those that can have a key labeled EXP, EEX, or perhaps EE that allows you to enter the exponent of 10. To enter a number such as 3×10^8, press the keys 3 EXP 8. To enter a number with a negative exponent, you must use the change-sign key, usually labeled +/− or CHS. To enter the number 5.2×10^{-3}, press the keys 5.2 EXP +/− 3. Try a few examples.

To read a number in scientific notation from a calculator, you must read the exponent separately. The number 3.1×10^{25} may appear in a calculator display as 3.1 25 or on some calculators as 3.1 10^{25}. Examine your calculator to determine how such numbers are displayed.

Astronomy Units and Constants

Astronomy, and science in general, is a way of learning about nature and understanding the universe. To test hypotheses about how nature works, scientists use observations of nature. The tables that follow contain some of the basic observations that support science's best understanding of the astronomical universe. Of course, these data are expressed in the form of numbers, not because science reduces all understanding to mere numbers, but because the struggle to understand nature is so demanding that science must use every valid means available. Quantitative thinking—reasoning mathematically—is one of the most powerful techniques ever invented by the human brain. Thus, these tables are not nature reduced to mere numbers but numbers supporting humanity's growing understanding of the natural world around us.

Table A-5 Astronomical Constants

Velocity of light *(c)*	$= 3.00 \times 10^8$ m/s
Gravitational constant *(G)*	$= 6.67 \times 10^{-11}$ m^3/s^2kg
Mass of H atom	$= 1.67 \times 10^{-27}$ kg
Mass of Earth $(M_\oplus)$	$= 5.97 \times 10^{24}$ kg
Earth equatorial radius $(R_\oplus)$	$= 6.38 \times 10^3$ km
Mass of Sun $(M_\odot)$	$= 1.99 \times 10^{30}$ kg
Radius of Sun $(R_\odot)$	$= 6.96 \times 10^8$ m
Solar luminosity $(L_\odot)$	$= 3.83 \times 10^{26}$ J/s
Mass of Moon	$= 7.35 \times 10^{22}$ kg
Radius of Moon	$= 1.74 \times 10^3$ km

Table A-6 Units Used in Astronomy

1 Angstrom (Å)	$= 10^{-8}$ cm
	$= 10^{-10}$ m
	= 10 nm
1 astronomical unit (AU)	$= 1.50 \times 10^{11}$ m
	$= 93.0 \times 10^6$ mi
1 light-year (ly)	$= 6.32 \times 10^4$ AU
	$= 9.46 \times 10^{15}$ m
	$= 5.88 \times 10^{12}$ mi
1 parsec (pc)	$= 2.06 \times 10^5$ AU
	$= 3.09 \times 10^{16}$ m
	= 3.26 ly
1 kiloparsec (kpc)	= 1000 pc
1 megaparsec (Mpc)	= 1,000,000 pc

Table A-7 Properties of Main-Sequence Stars

Spectral Type	Absolute Visual Magnitude (M_v)	Luminosity[1,2]	Temp. (K)	λ_{max} (nm)	Mass[1]	Radius[1]	Average Density (g/cm^3)
O6	−5.7	600,000	38,000	76	40	18	0.01
B0	−4.0	40,000	30,000	97	18	7.4	0.06
B5	−1.2	800	16,000	181	5.0	3.8	0.1
A0	0.7	30	10,000	290	2.1	1.8	0.5
A5	2.0	13	8600	337	1.7	1.6	0.6
F0	2.7	4.1	7200	402	1.5	1.3	1.0
F5	3.5	2.3	6500	446	1.3	1.2	1.1
G0	4.4	1.3	5900	491	1.1	1.1	1.2
G2	4.8	1.0	5800	500	1.0	1.0	1.4
G5	5.1	0.7	5600	518	0.9	0.9	1.7
K0	5.9	0.4	5200	557	0.8	0.8	2.2
K5	7.4	0.2	4400	659	0.7	0.7	2.9
M0	8.8	0.1	3900	743	0.5	0.6	3.2
M5	16.3	0.01	3100	935	0.2	0.3	10

[1]Luminosity, mass, and radius are expressed in terms of the Sun's luminosity, mass, and radius. The Sun's luminosity, mass, and radius are given in Table A-5.
[2]Luminosity is computed from radius and temperature.

Table A-8 The 15 Brightest Stars

Star	Name	Apparent Visual Magnitude (m_v)	Distance (pc)	Absolute Visual Magnitude (M_v)	Spectral Type
	Sun	−26.74		4.8	G2 V
α CMa	Sirius	−1.47	2.6	1.4	A1 V
α Car	Canopus	−0.72	96	−5.6	F0 II
α Cen	Rigil Kentaurus	−0.29	1.3	4.1	G2 V
α Boo	Arcturus	−0.04	11	−0.3	K2 III
α Lyr	Vega	0.03	7.8	0.5	A0 V
α Aur	Capella	0.08	13	−0.5	G8 III
β Ori	Rigel	0.12	240	−6.8	B8 Iab
α CMi	Procyon	0.34	3.5	2.6	F5 IV-V
α Eri	Achernar	0.43	44	−2.8	B3 V
α Ori	Betelguese	0.58	200	−5.9	M2 Iab
β Cen	Hadar	0.60	120	−4.8	B1 III
α Aql	Altair	0.77	5.1	2.2	A7 V
α Cru	Acrux	0.81	98	−4.2	B0 IV
α Tau	Aldebaran	0.85	20	−0.7	K5 III

Table A-9 The 15 Nearest Stars

Name	Distance (ly)	Distance (pc)	Apparent Visual Magnitude (m_v)	Absolute Visual Magnitude (M_v)	Spectral Type
Sun			−26.7	4.8	G2
Proxima Cen	4.2	1.3	11.0	15.5	M6
α Cen A	4.4	1.3	0.0	4.4	G2
α Cen B	4.4	1.3	1.3	5.7	K1
Barnard's Star	5.9	1.8	9.6	13.2	M4
Wolf 359	7.8	2.4	13.5	16.6	M6
Lalande 21185	8.3	2.5	7.5	10.4	M2
α CMa A (Sirius A)	8.6	2.6	−1.5	1.4	A1
α CMa B (Sirius B)	8.6	2.6	8.4	11.3	White dwarf
Luyten 726-8A	8.7	2.7	12.6	15.5	M5
Luyten 726-8B	8.7	2.7	13.1	15.9	M6
Ross 154	9.7	3.0	10.4	13.1	M4
Ross 248	10.3	3.2	12.3	14.8	M6
ϵ Eri	10.5	3.2	3.7	6.2	K2
Lacaille 9352	10.7	3.3	7.3	9.8	M1

Table A-10 Properties of the Planets

ORBITAL PROPERTIES

Planet	Semimajor Axis (a) (AU)	Semimajor Axis (a) (10^6 km)	Orbital Period (P) (y)	Orbital Period (P) (days)	Average Orbital Velocity (km/s)	Orbital Eccentricity	Inclination to Ecliptic
Mercury	0.387	57.9	0.241	88.0	47.9	0.206	7.0°
Venus	0.723	108	0.615	224.7	35.0	0.007	3.4°
Earth	1.00*	150	1.00*	365.3	29.8	0.017	0°*
Mars	1.52	228	1.88	687.0	24.1	0.093	1.8°
Jupiter	5.20	779	11.9	4333	13.1	0.048	1.3°
Saturn	9.55	1429	29.5	10,759	9.6	0.056	2.5°
Uranus	19.22	2875	84.0	30,688	6.8	0.047	0.8°
Neptune	30.10	4504	164.8	60,182	5.4	0.009	1.8°

*By definition.

PHYSICAL PROPERTIES (Earth = ⊕)

Planet	Equatorial Radius (km)	Equatorial Radius (⊕ = 1)	Mass (⊕ = 1)*	Average Density (g/cm^3)	Surface Gravity (⊕ = 1)	Escape Velocity (km/s)	Sidereal Period of Rotation	Inclination of Equator to Orbit
Mercury	2440	0.383	0.055	5.43	0.38	4.3	58.6 d	0.0°
Venus	6052	0.949	5.24	5.20	0.90	10.4	243.0 d	177.3°
Earth	6378	1.00	1.000	5.51	1.00	11.2	23.93 h	23.4°
Mars	3396	0.533	0.107	3.93	0.38	5.0	24.62 h	25.2°
Jupiter	71,492	11.2	318	1.33	2.53	59.5	9.92 h	3.1°
Saturn	60,268	9.45	95.2	0.69	1.06	35.5	10.57 h	26.7°
Uranus	25,559	4.01	14.5	1.27	0.89	21.3	17.23 h	97.8°
Neptune	24,764	3.88	17.1	1.64	1.14	23.5	16.11 h	28.3°

*Earth's mass = 5.97 x 10^{24} kg

Table A-11 Principal Satellites of the Solar System

Primary	Satellite	Radius (km)	Distance from Primary (10^3 km)	Orbital Period (days)	Orbital Eccentricity	Orbital Inclination**
Earth	Moon	1738	384.4	27.32	0.055	18.3°
Mars	Phobos	14 × 12 × 10	9.4	0.32	0.018	1.0°
	Deimos	8 × 6 × 5	23.5	1.26	0.002	2.8°
Jupiter	Amalthea	135 × 100 × 78	182	0.50	0.003	0.4°
	Io	1820	422	1.77	0.000	0.3°
	Europa	1560	671	3.55	0.000	0.5°
	Ganymede	2630	1071	7.16	0.002	0.2°
	Callisto	2410	1884	16.69	0.008	0.2°
	Himalia	~85*	11,470	250.6	0.158	27.6°
Saturn	Janus	110 × 80 × 100	151.5	0.70	0.007	0.1°
	Mimas	196	185.5	0.94	0.020	1.5°
	Enceladus	260	238.0	1.37	0.004	0.0°
	Tethys	530	294.7	1.89	0.000	1.1°
	Dione	560	377	2.74	0.002	0.0°
	Rhea	765	527	4.52	0.001	0.4°
	Titan	2575	1222	15.94	0.029	0.3°
	Hyperion	205 × 130 × 110	1484	21.28	0.104	~0.5°
	Iapetus	720	3562	79.33	0.028	14.7°
	Phoebe	105	12,930	550.4	0.163	150°
Uranus	Miranda	235	129.9	1.41	0.017	3.4°
	Ariel	580	190.9	2.52	0.003	0°
	Umbriel	585	266.0	4.14	0.003	0°
	Titania	805	436.3	8.71	0.002	0°
	Oberon	790	583.4	13.46	0.001	0°
Neptune	Proteus	205	117.6	1.12	~0	~0°
	Triton	1355	354.59	5.88	0.00	160°
	Nereid	170	5588.6	360.12	0.76	27.7°
Pluto	Charon	604	19.6	6.39	~0	0°
	Nix	25 × 18 × 16	48.7	24.85	0.002	0.1°
	Hydra	32 × 22 × 12	64.7	38.20	0.006	0.2°

**Relative to primary's equator
*The ~ symbol means "approximately."

Table A-12 Meteor Showers

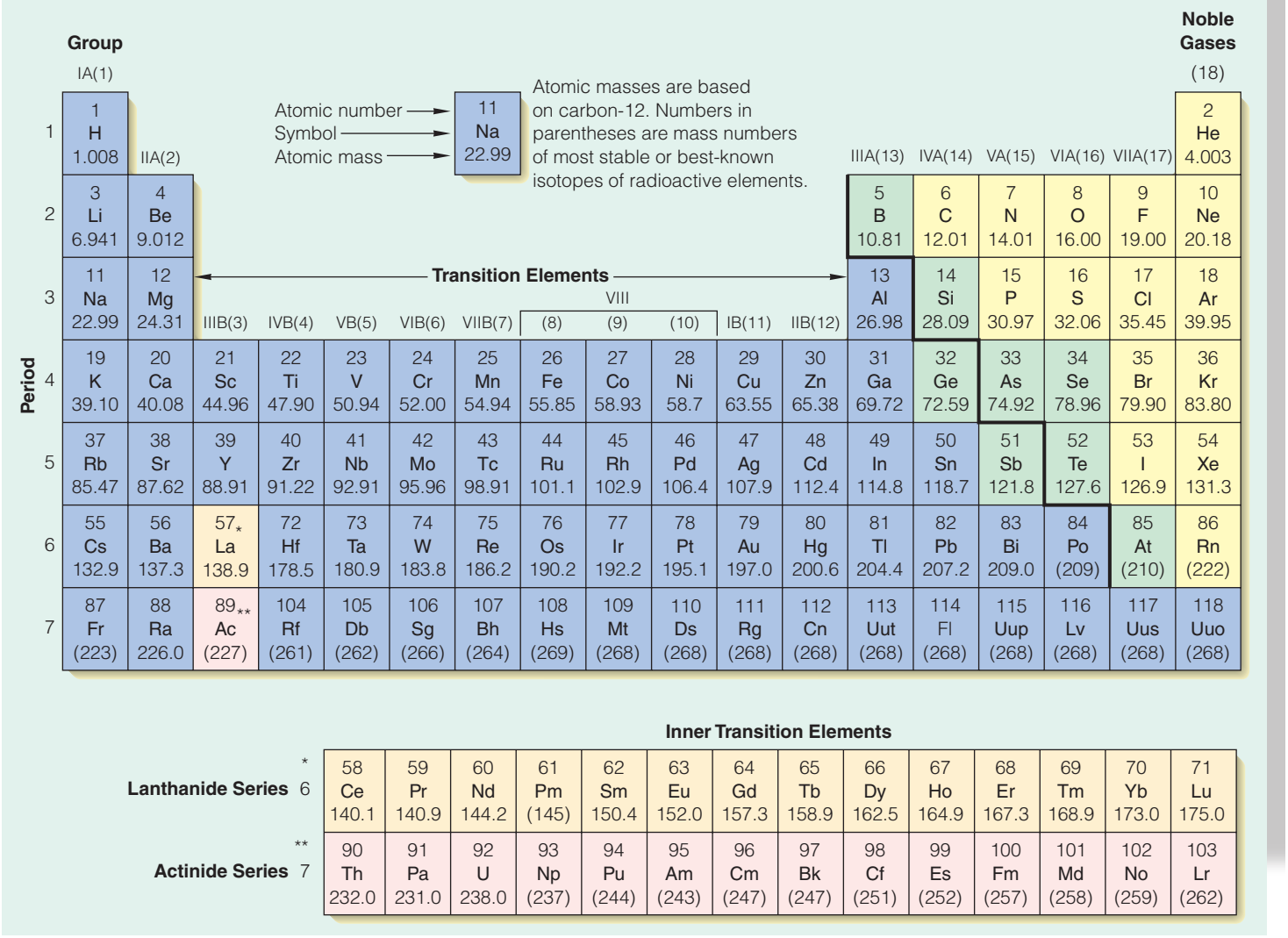
Shower	Dates	Rate per Hour	Radiant position		Associated Comet
			Right Ascension	Declination	
Quadrantids	Jan. 2–4	30	15h 24m	+50°	2003 EH1*
Lyrids	April 20–22	8	18h 08m	+33°	Thatcher
η Aquarids	May 2–7	10	22h 32m	−01°	Halley
S. δ Aquarids	July 26–31	15	22h 40m	−10°	?
Perseids	Aug. 10–14	40	03h 12m	+58°	Swift-Tuttle
Orionids	Oct. 18–23	15	06h 20m	+16°	Halley
S. Taurids	Nov. 1–7	8	03h 40m	+15°	Encke
Leonids	Nov. 14–19	6	10h 16m	+22°	Tempel-Tuttle
Geminids	Dec. 10–13	50	07h 24m	+33°	Phaethon*

*Source object appears asteroid-like.

Table A-13 The Greek Alphabet

A, α	alpha	H, η	eta	N, ν	nu	T, τ	tau
B, β	beta	Θ, θ	theta	Ξ, ξ	xi	Υ, υ	upsilon
Γ, γ	gamma	I, ι	iota	O, o	omicron	Φ, ϕ	phi
Δ, δ	delta	K, κ	kappa	Π, π	pi	X, χ	chi
E, ϵ	epsilon	Λ, λ	lambda	P, ρ	rho	Ψ, ψ	psi
Z, ζ	zeta	M, μ	mu	Σ, σ	sigma	Ω, ω	omega

Table A-14 Periodic Table of the Elements



The Elements and Their Symbols

Actinium	Ac	Cesium	Cs	Germanium	Ge	Manganese	Mn	Potassium	K	Strontium	Sr
Aluminum	Al	Chlorine	Cl	Gold	Au	Meitnerium	Mt	Praseodymium	Pr	Sulfur	S
Americium	Am	Chromium	Cr	Hafnium	Hf	Mendelevium	Md	Promethium	Pm	Tantalum	Ta
Antimony	Sb	Cobalt	Co	Hassium	Hs	Mercury	Hg	Protactinium	Pa	Technetium	Tc
Argon	Ar	Copernicium	Cn	Helium	He	Molybdenum	Mo	Radium	Ra	Tellurium	Te
Arsenic	As	Copper	Cu	Holmium	Ho	Neodymium	Nd	Radon	Rn	Terbium	Tb
Astatine	At	Curium	Cm	Hydrogen	H	Neon	Ne	Rhenium	Re	Thallium	Tl
Barium	Ba	Darmstadtium	Ds	Indium	In	Neptunium	Np	Rhodium	Rh	Thorium	Th
Berkelium	Bk	Dubnium	Db	Iodine	I	Nickel	Ni	Roentgenium	Rg	Thulium	Tm
Beryllium	Be	Dysprosium	Dy	Iridium	Ir	Niobium	Nb	Rubidium	Rb	Tin	Sn
Bismuth	Bi	Einsteinium	Es	Iron	Fe	Nitrogen	N	Ruthenium	Ru	Titanium	Ti
Bohrium	Bh	Erbium	Er	Krypton	Kr	Nobelium	No	Rutherfordium	Rf	Tungsten	W
Boron	B	Europium	Eu	Lanthanum	La	Osmium	Os	Samarium	Sm	Uranium	U
Bromine	Br	Fermium	Fm	Lawrencium	Lr	Oxygen	O	Scandium	Sc	Vanadium	V
Cadmium	Cd	Flerovium	Fl	Lead	Pb	Palladium	Pd	Seaborgium	Sg	Xenon	Xe
Calcium	Ca	Fluorine	F	Lithium	Li	Phosphorous	P	Selenium	Se	Ytterbium	Yb
Californium	Cf	Francium	Fr	Livermorium	Lv	Platinum	Pt	Silicon	Si	Yttrium	Y
Carbon	C	Gadolinium	Gd	Lutetium	Lu	Plutonium	Pu	Silver	Ag	Zinc	Zn
Cerium	Ce	Gallium	Ga	Magnesium	Mg	Polonium	Po	Sodium	Na	Zirconium	Zr

Observing the Sky

Observing the sky with the unaided eye is as important to modern astronomy as picking up pretty pebbles is to modern geology. The sky is a natural wonder unimaginably bigger than the Grand Canyon, the Rocky Mountains, or any other site that tourists visit every year. To neglect the beauty of the sky is equivalent to geologists neglecting the beauty of the minerals they study. This supplement is meant to act as a tourist's guide to the sky. You analyzed the universe in the textbook's chapters; here you can admire it.

The brighter stars in the sky are visible even from the centers of cities with their air and light pollution. But in the countryside, only a few miles beyond the cities, the night sky is a velvety blackness strewn with thousands of glittering stars. From a wilderness location, far from the city's glare, and especially from high mountains, the night sky is spectacular.

Using Star Charts

The constellations are a fascinating cultural heritage of our planet, but they are sometimes a bit difficult to learn because of Earth's motion. The constellations above the horizon change with the time of night and the seasons.

Because Earth rotates eastward, the sky appears to rotate westward around Earth. A constellation visible overhead soon after sunset will appear to move westward, and in a few hours it will disappear below the horizon. Other constellations will rise in the east, so the sky changes gradually through the night.

In addition, Earth's orbital motion makes the Sun appear to move eastward among the stars. Each day the Sun moves about twice its own diameter, about one degree, eastward along the ecliptic. Consequently, each night at sunset, the constellations are about one degree farther toward the west.

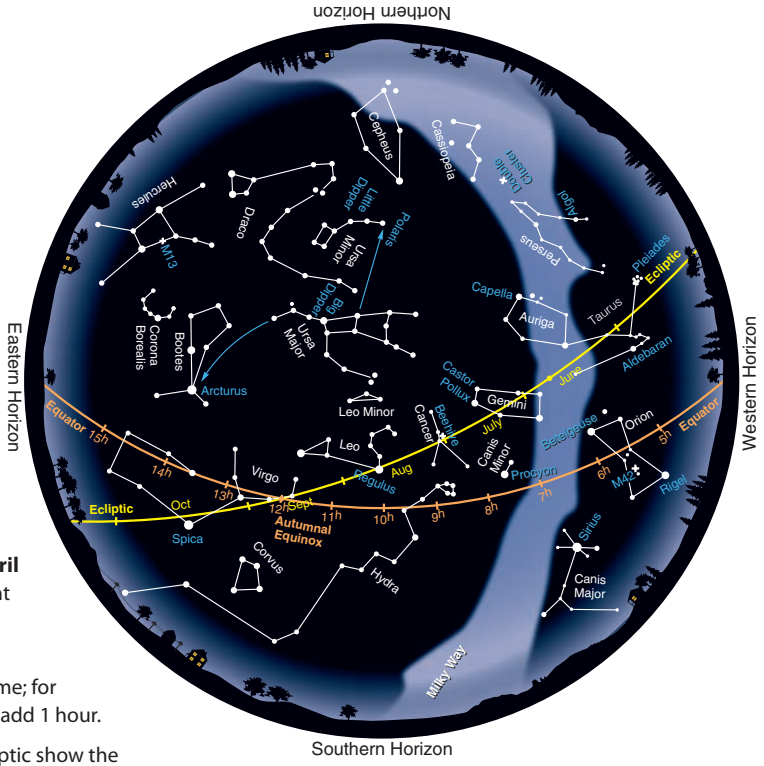
Orion, for instance, is visible in the evening sky in January, but, as the days pass, the Sun moves closer to Orion. By March, Orion is difficult to see in the western sky soon after sunset. By June, the Sun is so close to Orion that the constellation sets with the Sun and is invisible. Not until late July is the Sun far enough past Orion for the constellation to become visible rising in the eastern sky just before dawn.

You need more than one star chart to map the visible sky because of the rotation and orbital motion of Earth. Which chart you select depends on the month and the time of night. The Northern Hemisphere star charts on the next two pages show the sky in three-month intervals as seen from latitudes typical for the United States and southern Europe.

To use the charts, select the appropriate chart and hold it overhead as shown in **Figure B-1**. If you face south, turn the chart until the words southern horizon are at the bottom of the chart. If you face other directions, turn the chart appropriately. Note that hours are in standard time; for daylight savings time (mid-March through early November in the United States) add one hour.

▲ **Figure B-1** To use the star charts in this book, select the appropriate chart for the season and time. Hold it overhead and turn it until the direction at the bottom of the chart is the same as the direction you are facing.

Star Charts



Northern Hemisphere Sky

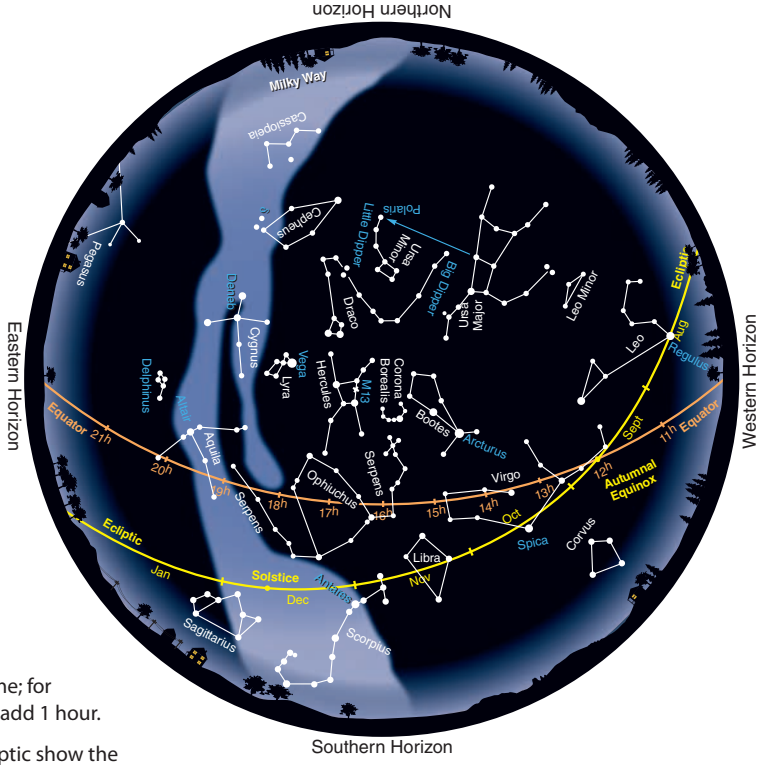
February–March–April

February	midnight
March	10 PM
April	8 PM

Times are Standard Time; for Daylight Saving Time, add 1 hour.

Months along the ecliptic show the location of the Sun during the year.

Numbers along the celestial equator show Right Ascension (celestial longitude).



Northern Hemisphere Sky

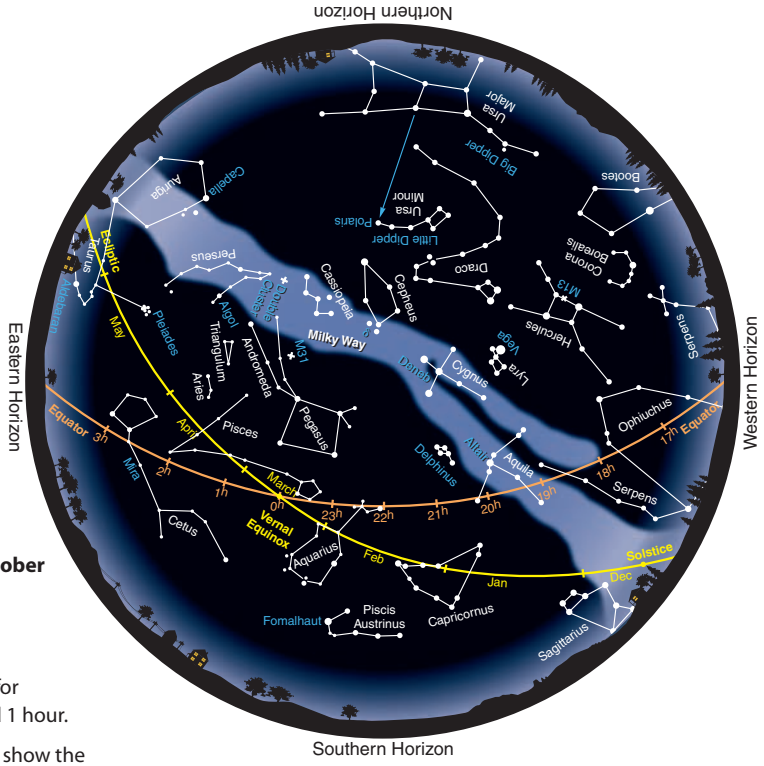
May–June–July

May	midnight
June	10 PM
July	8 PM

Times are standard time; for Daylight Saving Time, add 1 hour.

Months along the ecliptic show the location of the Sun during the year.

Numbers along the celestial equator show Right Ascension (celestial longitude).



Northern Hemisphere Sky

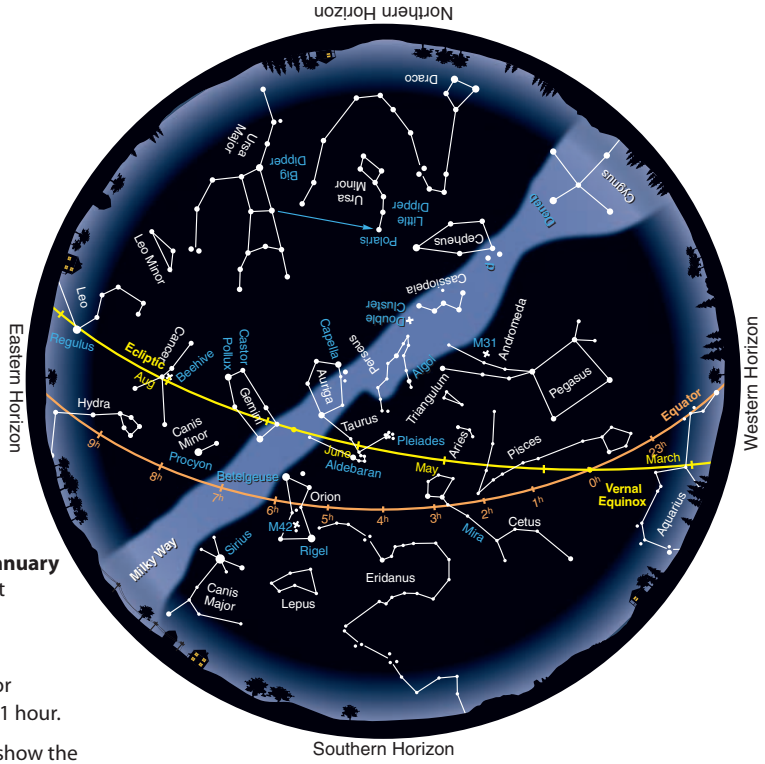
August–September–October

August	midnight
September	10 PM
October	8 PM

Times are standard time; for Daylight Saving Time, add 1 hour.

Months along the ecliptic show the location of the Sun during the year.

Numbers along the celestial equator show Right Ascension (celestial longitude).



Northern Hemisphere Sky

November–December–January

November	midnight
December	10 PM
January	8 PM

Times are standard time; for Daylight Saving Time, add 1 hour.

Months along the ecliptic show the location of the Sun during the year.

Numbers along the celestial equator show Right Ascension (celestial longitude).

ANSWERS TO EVEN-NUMBERED PROBLEMS

Chapter 1
2. 2160 miles; **4.** 1×10^{11} stars; **6.** 1.1×10^{8} km; **8.** about 1.2 seconds; **10.** 640 ly; approx. beginning of the Renaissance in Europe, beginning of the Ming Dynasty in China; Betelgeuse is nearby compared with most stars in the Universe, but farther than most stars visible with an unaided eye; **12.** about 2.4×10^{22} m

Chapter 2
2. about 2.2 mags; **4.** about 0.75 mag; Star A is brighter; **6.** +3.8; both Star A and Star B can be seen with unaided eyes; **8.** about 170; **10.** a factor of about 1000 (7.5 mags); **12.** 40 degrees; 40 degrees; **14.** about 1.6; about 3.8; the largest climate variations in Figure 2-11 have a period of about 100,000 years, the time scale for variations in Earth's orbit shape

Chapter 3
2. new; full; first quarter; third quarter; **4.** about ½; ½; ½; **6.** about 12 degrees per day; **8.** about 2 to 3 times larger; **10.** 207 degrees; note, in this situation Earth would actually be inside the Sun and the small-angle approximation would not be accurate; **12.** 4/10, or 40 percent; **14.** (a) The Moon won't be full until about October 18; (b) The Moon will no longer be near that node of its orbit; **16.** August 12, 2026 [July 10, 1972 + 3 × (6585 ⅓ days)]. Note that you must take into account the number of leap days in the interval to get the right answer; **18.** S S S S S S N N S S S S N N S S S S S S; an educated guess would be N N

Chapter 4
2. retrograde motion: Mercury, Venus, Earth, Jupiter, Saturn, Uranus, and Neptune; never seen as crescents: Jupiter, Saturn, Uranus, and Neptune; **4.** 2.8 yr; **6.** 19.2 AU; **8.** more; faster; 2.8 times; 30 AU, Saturn and Neptune; **10.** new; full; Venus orbits the Sun, which is the light source

Chapter 5
2. 100 times weaker; 400 times weaker; **4.** increase, by a factor of 4; **6.** 19.6 m/s; 39.2 m/s; **8.** about 3070 m/s (3.07 km/s); about 86,700 s (slightly more than 24 hrs); **10.** The cannonball would move in an elliptical orbit with Earth's center at one focus of the ellipse; **12.** 6320 s (about 1 hr 45 min); **14.** about 1.0×10^{-4} m/s (0.1 mm/s); a fly could easily escape

Chapter 6
2. 3 m; **4.** less; red and blue, respectively; blue; 1.75 times as much; **6.** Each Keck telescope has an LGP 1.6 million times greater than a human eye; **8.** Telescope A (diameter = 60 in. = 1.52 m), 0.075 arc seconds; Telescope B, 2.8 arc seconds; Telescope A has better resolving power; it can resolve smaller angular size; **10.** 0.46 arc seconds; the two stars would be resolved as separate objects; **12.** The Palomar telescope has about 2.1 times better (smaller) resolving power than *HST,* considering only diffraction; *HST* is in space and is not subject to atmospheric seeing; **14.** Resolving power requires a diameter of 1.13 meters; magnifying power requires a primary focal length 250 times larger than the eyepiece focal length; light-gathering power will be about 20,000 times that of the dark-adapted human eye

Chapter 7
2. 5.8×10^{4} nm, infrared, object in the outer Solar System; 5.8×10^{3} nm, infrared, object in the inner Solar System; 580 nm, visible, medium-temperature star; 58 nm, ultraviolet, hot star; Wien's law: wavelength of blackbody maximum intensity is inversely proportional to temperature; **4.** 1450 K; **6.** about 1.9 times more; **8.** 250 nm; **10.** about 120 km/s; receding; redshift; **12.** about 0.65 nm

Chapter 8
2. core about 1/3 of the Sun's radius or 1/27 (approximately 4%) of the volume; the radiative zone about 8/27 (approximately 30%) of the volume, and the convective zone about 70%; **4.** about 6×10^{-3} arc seconds; no, that is much smaller than *HST*'s resolution; **6.** 6.3×10^{7} watts; increases by a factor of 16; **8.** 656.299 nm; blueshift; rising; **10.** average rotation period is about 28.9 days; **12.** about 6×10^{14} joules; **14.** 2.5×10^{9} megatons; **16.** 4.3×10^{9} kg

Chapter 9
2. B = 16 ly, C = 7.9×10^{-5} ly; B, A, C; B and A could be stars, C could be Jupiter; **4.** 62 pc; +2.0; **6.** Star A is intrinsically brighter, by a factor of 16; **8.** infrared; deep red; it is a brown dwarf, spectral type indicates temperature too low to be a star; **10.** assuming M_v is approximately –0.7, distance is about 14 pc; class III (giant); **12.** assuming M_v is approx. –5, distance is about 16,000 pc; **14.** 3.87×10^{26} watts using 1 AU = 1.50×10^{13} m; less than 1 percent different from the Celestial Profile value of 3.84×10^{26} watts; **16.** 1.3 $M_\odot$; **18.** 4.26×10^{8} km (2.84 AU), 2.13×10^{8} km (1.42 AU); 0.678 AU; 40.6 $M_\odot$; 13.5 and 27.1 $M_\odot$; **20.** about 47 times more; they could be Vega and the Sun

Chapter 10
2. about 6×10^{4} nm (60 microns); **4.** about 1/10,000; **6.** 1420 MHz; UHF; **8.** about 85 $M_\odot$; **10.** 0.29 to 2.9 nm

Chapter 11
2. 120 arc seconds; **4.** 500 stars; **6.** 1.9×10^{4} nm (19 microns); infrared; **8.** about 3.2 times the Sun's radius; **10.** T must be greater than 31,800 K; about O8 or hotter; the star is type O6 (Fig. 2, Concept Art: Star Formation in the Orion Nebula); **12.** CNO cycle and proton–proton chain both release energy equivalent to the mass difference between 1 He nucleus and 4 H nuclei, according to $E = mc^2$; **14.** 72 nm; shorter than the hydrogen ionization threshold wavelength (photons will have more than enough energy)

Chapter 12
2. Almost all luminosity is produced within the inner 30 percent of the Sun's radius; about 20,000 times denser; about 3000 times hotter; **4.** Star B, by a factor of about 57,000; **6.** Star A, about 4×10^{-6} $t_\odot$ (solar lifetimes); Star B, about 600 $t_\odot$; Star A, about 4×10^{7} $L_\odot$; Star B, about 1×10^{-4} $L_\odot$; the Sun's lifetime is closer to that of the low-mass star than the high-mass star; **8.** about 1×10^{6} times less than present, or about 1.4×10^{-6} g/cm^3; **10.** about $(1/800)^3 = 2 \times 10^{-9}$; **12.** about 3 pc; **14.** average m_v is about +4.0 and M_v is about –3.4, so distance is about 300 pc; **16.** 183 seconds (about 3 minutes)

Chapter 13
2. about 6×10^{44} joules; **4.** m_v would be about –22.7, much brighter than a full moon; **6.** about 1.8 ly; **8.** about 140 pc; **10.** about 940 years ago, in approximately the year 1075, close to the year 1054 in which the supernova was actually observed; **12.** about 2400 pc; **14.** about 170,000 years ago; the neutrinos must have left before the light

Chapter 14
2. about 7.8 km; **4.** about 7×10^{17} kg/m^3 (7×10^{14} g/cm^3); about 3×10^{18} kg/m^3 (3×10^{15} g/cm^3); **6.** about 2×10^{8} km/s; about 3×10^{8} km/s (the speed of light!); **8.** about 2.9 nm; X-ray; **10.** a = 0.013 AU, which is 2.8 $R_\odot$; the text states (Section 14-1e) "roughly equal to the radius of our Sun"; 1 $R_\odot$ is probably the minimum distance in the elliptical orbit (Figure 14-6b); **12.** 970 km/s; **14.** in both cases, the speed of light, by definition of the Schwarzschild radius; **16.** 0.302 AU

Chapter 15
2. about 17 percent; **4.** about 4×10^{6} yr; **6.** for M_v about +0.5, d is about 8000 pc (8 kpc); **8.** about 21,000 pc (21 kpc); **10.** about 8.8×10^{10} $M_\odot$ (88 billion solar masses); **12.** hotter than about 1450 K; **14.** about +19.4; yes, observable by *HST*; +49.4; no, not observable by *HST*

Chapter 16
2. The a and b axes are equal; axis c is unobservable so you don't know the 3-D shape of the galaxy; **4.** for m_v about +0.4, d is about 330 pc; **6.** about 5.8 kpc; **8.** about 29 Mpc (29 million pc); **10.** about 530 km/s/Mpc; **12.** about 1.6×10^{8} yr (160 million years); **14.** 3.3×10^{9} yr (3.3 billion years); **16.** 4.5×10^{41} kg (2.3×10^{11} $M_\odot$)

Chapter 17
2. about 7.8×10^{6} yr (7.8 million years); **4.** 0.024 pc; **6.** about −28.5; **8.** about +16.1; **10.** 2.9 nm; X-ray; **12.** $\Delta\lambda$ = 2916.6 nm; the unshifted line is in the visual band, the shifted line is in the infrared; this quasar is much farther from Earth than 3C 273, which has a redshift of 0.158

Chapter 18
2. 0.203 c; particles in physics research accelerators can go this speed and faster; it is not an actual speed but rather the rate at which space is stretching between Earth and the Hydra galaxy cluster; **4.** The nearer raisin is moving away less quickly than the farther raisin. The rate of recession is directly proportional to the distance; **6.** for T = 3000 K, λ_{max} = 970 nm, in the near-infrared; deep red; **8.** 1.6×10^{-27} kg/m^3 (which is almost exactly one H atom per cubic meter); open universe; **10.** 76 km/s/Mpc (close to the modern value of 70 km/s/Mpc); **12.** about 240 million years

Chapter 19
2. 8.8 cm; 0.088 cm (0.88 mm); **4.** 10,000 generations; **6.** 380 km (240 mi) at average closest approach, d = 0.52 AU; **8.** 44,000 years; **10.** no definitely correct answer; perhaps somewhere between 2×10^{-4} and 1×10^{8}, the pessimistic and optimistic values in the table

GLOSSARY

absolute age An age determined in years, as from radioactive dating (see also *relative age*).

absolute visual magnitude (M_V) Intrinsic brightness of a star; the apparent visual magnitude the star would have if it were 10 pc away.

absolute zero The lowest possible temperature; the temperature at which the particles in a material, atoms or molecules, contain no energy of motion that can be extracted from the body.

absorption line A dark line in a spectrum; produced by the absence of photons absorbed by atoms or molecules.

absorption (dark-line) spectrum A spectrum that contains absorption lines.

acceleration A change in a velocity; a change in either speed or direction. (See *velocity*.)

acceleration of gravity A measure of the strength of gravity at a planet's surface.

accretion The sticking together of solid particles to produce a larger particle.

accretion disk The whirling disk of gas that forms around a compact object such as a white dwarf, neutron star, or black hole as matter is drawn in.

achondrite Stony meteorite containing no chondrules or volatiles.

achromatic lens A telescope lens composed of two lenses ground from different kinds of glass and designed to bring two selected colors to the same focus and correct for chromatic aberration.

active galactic nucleus (AGN) The central energy source of an active galaxy.

active galaxy A galaxy that is a source of excess radiation, usually radio waves, X-rays, gamma-rays, or some combination.

active optics Optical elements whose position or shape is continuously controlled by computers.

active region An area on the Sun where sunspots, prominences, flares, and the like occur.

adaptive optics Computer-controlled telescope mirrors that can at least partially compensate for seeing.

albedo The fraction of the light hitting an object that is reflected.

alt-azimuth mount A telescope mounting capable of motion parallel to and perpendicular to the horizon.

Amazonian period On Mars, the geological era from about 3 billion years ago to the present marked by low-level cratering, wind erosion, and small amounts of water seeping from subsurface ice.

amino acid One of the carbon-chain molecules that are the building blocks of protein.

angstrom (Å) A unit of distance; $1\ \text{Å} = 1 \times 10^{-10}$ m; often used to measure the wavelength of light.

angular diameter A measure of the size of an object in the sky; numerically equal to the angle in degrees between two lines extending from the observer's eye to opposite edges of the object.

angular distance A measure of the separation between two objects in the sky; numerically equal to the angle in degrees between two lines extending from the observer's eye to the two objects.

angular momentum The tendency of a rotating body to continue rotating; mathematically, the product of mass, velocity, and radius.

annular eclipse A solar eclipse in which the solar photosphere appears around the edge of the Moon in a bright ring, or annulus. The corona, chromosphere, and prominences cannot be seen.

anorthosite Rock of aluminum and calcium silicates found in the lunar highlands.

anticoded Describes a message designed to be understood by a recipient about whom the sender knows little or nothing, for example an interstellar broadcast aimed at possible inhabitants of another planet.

antimatter Matter composed of antiparticles, which on colliding with a matching particle of normal matter annihilate and convert the mass of both particles into energy. The antiproton is the antiparticle of the proton, and the positron is the antiparticle of the electron.

aphelion The orbital point of greatest distance from the Sun.

apogee The orbital point of greatest distance from Earth.

Apollo–Amor object Asteroid whose orbit crosses that of Earth (Apollo) and Mars (Amor).

apparent visual magnitude, m_V The brightness of a star as seen by human eyes on Earth.

arc minute An angular measure; each degree is divided into 60 arc minutes.

arc second An angular measure; each arc minute is divided into 60 arc seconds.

archaea A biological kingdom of microorganisms similar to, but distinct from, bacteria, with characteristics most closely resembling those inferred for the common ancestor of all present-day Earth life.

archaeoastronomy The study of the astronomy of ancient cultures.

array detector A grid of photosensitive detectors for the purpose of recording images. Commercial still and video digital cameras contain CCD array detectors. See also *charge-coupled device*.

association Group of widely scattered stars (10 to 1000) moving together through space; not gravitationally bound into a cluster.

asterism A named group of stars not identified as a constellation, for example, the Big Dipper.

asteroid Small, rocky world; most asteroids lie between Mars and Jupiter in the asteroid belt.

astrobiology The field of study involving searches for life on other worlds and investigation of possible habitats for such life. Also known as "exobiology."

astronomical unit (AU) Average distance from Earth to the Sun; 1.5×10^8 km, or 93×10^6 miles.

atmospheric window Wavelength region in which Earth's atmosphere is transparent—at visual, infrared, and radio wavelengths.

aurora The glowing light display that results when a planet's magnetic field guides charged particles toward the north and south magnetic poles, where they strike the upper atmosphere and excite atoms to emit photons.

autumnal equinox The point on the celestial sphere where the Sun crosses the celestial equator going southward. Also, the time when the Sun reaches this point and autumn begins in the Northern Hemisphere—about September 22.

Babcock model A model of the Sun's magnetic cycle in which the differential rotation of the Sun winds up and tangles the solar magnetic field in a 22-year cycle. This is thought to be responsible for the 11-year sunspot cycle.

Balmer series Spectral lines in the visible and near-ultraviolet spectrum of hydrogen produced by transitions whose lowest orbit is the second.

barred spiral galaxy A spiral galaxy with an elongated nucleus resembling a bar from which the arms originate.

basalt Dark, igneous rock characteristic of solidified lava.

belt–zone circulation The atmospheric circulation typical of Jovian planets. Dark belts and bright zones encircle the planet parallel to its equator.

Big Bang The theory that the Universe began with a violent explosion from which the expanding Universe of galaxies eventually formed.

big rip The possible fate of the Universe if dark energy increases rapidly and the expansion of space-time pulls galaxies, stars, and ultimately atoms apart.

binary star One of a pair of stars that orbit around their common center of mass.

binding energy The energy needed to pull an electron away from its atom.

biological evolution The combined effect of variation and natural selection resulting in new species arising and existing species adapting to the environment or becoming extinct. Also called "Darwinian evolution."

bipolar flow Oppositely directed jets of gas ejected by some protostellar objects.

birth line In the H–R diagram, the line above the main sequence where protostars first become visible.

black dwarf The end state of a white dwarf that has cooled to low temperature.

black hole A mass that has collapsed to such a small volume that its gravity prevents the escape of all radiation; also, the volume of space from which radiation may not escape.

blackbody radiation Radiation emitted by a hypothetical perfect radiator; the spectrum is continuous, and the wavelength of maximum emission depends only on the body's temperature.

blue dwarf galaxy Small irregular clouds forming massive hot stars at a rapid rate. These are observed only at large look-back times, that is, soon after the Big Bang. They may be the units from which larger galaxies such as the Milky Way were later assembled.

blueshift The shortening of the wavelengths of light observed when the source and observer are approaching each other.

Bok globule Small, dark, and dense interstellar cloud only about 1 ly in diameter that contains 10 to 1000 solar masses of gas and dust; thought to be related to star formation.

bottom-up hypothesis The conjecture that the Milky Way Galaxy and other large galaxies formed mostly by collisions and combination of smaller galaxies and star clusters. See also *monolithic collapse* or *top-down hypothesis*.

bow shock The boundary between the undisturbed solar wind and the region being deflected around a planet or comet.

breccia A rock composed of fragments of earlier rocks bonded together.

brown dwarf A very cool, low-luminosity star whose mass is not sufficient to ignite nuclear fusion.

butterfly diagram See *Maunder butterfly diagram*.

CAI Calcium–aluminum-rich inclusions found in some meteorites.

calibrate To make the observations of reference objects, checks on instrument performance, calculations of units conversions, and so on, needed to completely understand measurements of unknown quantities.

Cambrian explosion The sudden appearance of complex life forms at the beginning of the Cambrian period 0.6 to 0.5 billion years ago. Cambrian rocks contain the oldest easily identifiable fossils.

carbon deflagration The process in which the carbon in a white dwarf is completely consumed by nuclear fusion, producing a type Ia supernova explosion.

carbonaceous chondrite Stony meteorite that contains both chondrules and volatiles. These may be the least altered remains of the solar nebula still present in the Solar System.

carbon–nitrogen–oxygen (CNO) cycle A series of nuclear reactions that use carbon as a catalyst to combine four hydrogen atoms to make one helium atom plus energy; effective in stars more massive than the Sun.

Cassegrain focus The optical design of a reflecting telescope in which the secondary mirror reflects light back down the tube through a hole in the center of the objective mirror.

catastrophic hypothesis Explanation for natural processes that depends on dramatic and unlikely events, such as the collision of two stars to produce our Solar System.

celestial equator The imaginary line around the sky directly above Earth's equator.

celestial sphere An imaginary sphere of very large radius surrounding Earth to which the planets, stars, Sun, and Moon seem to be attached.

centaur An outer Solar System body with an orbit entirely within the region of the Jovian planets, for example Chiron, that orbits between Saturn and Uranus.

center of mass The balance point of a body or system of bodies.

central bulge The spheroidal cloud of stars at the center of most spiral galaxies, including our Milky Way Galaxy.

Cepheid variable star Variable star with a period of 1 to 60 days; the period of variation is related to luminosity.

Chandrasekhar limit The maximum mass of a white dwarf, about 1.4 solar masses; a white dwarf of greater mass cannot support itself and will collapse.

charge-coupled device (CCD) An electronic device consisting of a large array of light-sensitive elements used to record very faint images.

chemical evolution The chemical process that led to the growth of complex molecules on the primitive Earth. This did not involve the reproduction of molecules.

Chicxulub The buried crater associated with the mass extinction event at the end of the age of dinosaurs, named after the town in the coastal region of Mexico's Yucatán peninsula near the center of the crater.

chondrite A stony meteorite that contains chondrules.

chondrule Round, glassy body in some stony meteorites; thought to have solidified very quickly from molten drops of silicate material.

chromatic aberration A distortion found in refracting telescopes because lenses focus different colors at slightly different distances. Images are consequently surrounded by color fringes.

chromosome One of the bodies in a cell that contains the DNA carrying genetic information.

chromosphere Bright gases just above the photosphere of the Sun.

circular velocity The velocity required to remain in a circular orbit about a body.

circumpolar constellation An constellations so close to the celestial pole that it never sets (or never rises) as seen from a given latitude.

closed orbit An orbit that returns to its starting point; a circular or elliptical orbit. (See *open orbit*.)

closed universe A model universe in which the average density is great enough to stop the expansion and make the Universe contract.

cluster method The method of determining the masses of galaxies based on the motions of galaxies in a cluster.

CNO cycle See *carbon–nitrogen–oxygen cycle*.

cocoon nebula The cloud of gas and dust around a contracting protostar that conceals it at visible wavelengths.

cold dark matter Invisible matter in the Universe composed of heavy, slow-moving particles such as WIMPs.

coma The glowing head of a comet.

comet One of the small, icy bodies that orbit the Sun and produce tails of gas and dust when they near the Sun.

compact object A star that has collapsed to form a white dwarf, neutron star, or black hole.

comparative planetology The study of planets by comparing the characteristics of different examples.

comparison spectrum A spectrum of known spectral lines used to identify unknown wavelengths in an object's spectrum.

composite volcano A volcano built up of layers of lava flows and ash falls. These are steep sided and typically associated with subduction zones.

condensation The growth of a particle by addition of material from surrounding gas, one atom or molecule at a time.

condensation sequence The sequence in which different materials condense from the solar nebula at increasing distances from the Sun.

constellation One of the stellar patterns identified by name, usually of mythological gods, people, animals, or objects; also, the region of the sky containing that star pattern.

continuous spectrum A spectrum in which there are no absorption or emission lines.

convection Circulation in a fluid driven by heat; hot material rises, and cool material sinks.

convective zone The region inside a star where energy is carried outward as rising hot gas and sinking cool gas.

cooling line A far-infrared spectral emission line that releases energy from the insides of contracting protostellar clouds, allowing rapid contraction to continue.

Copernican Principle The idea that Earth should not be assumed to be in a special location or to be otherwise unique.

corona The faint outer atmosphere of the Sun; composed of low-density, very hot, ionized gas. On Venus, round network of fractures and ridges up to 1000 km in diameter, caused by the intrusion of magma below the crust.

coronagraph A telescope designed to photograph the inner corona of the Sun.

coronal gas Extremely high-temperature, low-density gas in the interstellar medium.

coronal hole An area of the solar surface that is dark at X-ray wavelengths; thought to be associated with divergent magnetic fields and the source of the solar wind.

coronal mass ejection (CME) Gas trapped in the Sun's magnetic field.

cosmic microwave background radiation Radiation from the hot clouds of the Big Bang explosion. Because of its large redshift, it appears to come from a body whose temperature is only 2.7 K.

cosmic ray A subatomic particle traveling at tremendous velocity that strikes Earth's atmosphere from space.

cosmological constant (λ) Einstein's constant that represents a repulsion in space to oppose gravity.

cosmological principle The assumption that any observer in any galaxy sees the same general features of the Universe.

cosmologist An astronomer or physicist whose research focuses on the overall properties of the Universe and its origin.

cosmology The study of the nature, origin, and evolution of the Universe.

Coulomb barrier The electrostatic force of repulsion between bodies of like charge; commonly applied to atomic nuclei.

Coulomb force The repulsive force between particles with like electrostatic charge.

critical density The average density of the Universe needed to make its curvature flat.

critical point The temperature and pressure at which the vapor and liquid phases of a material have the same density.

C-type asteroid A type of asteroid common in the outer asteroid belt, with very low reflectivity and grayish color, probably composed of carbonaceous material.

dark age The period of a few hundred million years during which the Universe expanded in darkness. Extends from soon after the Big Bang glow faded into the infrared to the formation of the first stars.

dark energy The energy of empty space that drives the acceleration of the expanding Universe.

dark matter Nonluminous material that is detected only by its gravitational influence.

dark nebula A nonluminous cloud of gas and dust visible because it blocks light from more distant stars and nebulae.

debris disk A disk of dust found by infrared observations around some stars. The dust is debris from collisions among asteroids, comets, and Kuiper Belt objects.

deferent In the Ptolemaic theory, the large circle around Earth along which the center of the epicycle moved.

degenerate matter Extremely high-density matter in which pressure no longer depends on temperature, quantum mechanical effects.

dense core Within an interstellar molecular cloud, an especially dense concentration of material that is possibly destined to become a star.

density The amount of matter per unit volume in a material; measured in grams per cubic centimeter, for example.

deuterium An isotope of hydrogen in which the nucleus contains a proton and a neutron.

diamond ring effect A momentary phenomenon seen during some total solar eclipses when the ring of the corona and a bright spot of photosphere resemble a large diamond set in a silvery ring.

differential rotation The rotation of a body in which different parts of the body have different periods of rotation; this is true of the Sun, the Jovian planets, and the disk of the galaxy.

differentiation The separation of planetary material according to density.

diffraction fringe Blurred fringe surrounding any image caused by the wave properties of light. Because of this, no image detail smaller than the fringe can be seen.

direct collapse The hypothetical process by which a Jovian planet might skip the accretion of a solid core, instead forming quickly and directly from the gases of the solar nebula.

disk component All material confined to the plane of the galaxy.

distance indicator Object of known brightness that astronomers use to find distance—for example, Cepheid variable stars and supernovae. Sometimes called a standard candle.

distance scale The combined calibration of distance indicators used by astronomers to find the distances to remote galaxies.

DNA (deoxyribonucleic acid) The long carbon-chain molecule that records information to govern the biological activity of the organism. DNA carries the genetic data passed to offspring.

Doppler effect The change in the wavelength of radiation relative radial motion of source and observer.

double-exhaust model The theory that double radio lobes are produced by pairs of jets emitted in opposite directions from the centers of active galaxies.

double-lobed radio galaxy A galaxy that emits radio energy from two regions (lobes) located on opposite sides of the galaxy.

Drake equation A formula for the number of communicative civilizations in our galaxy.

dust tail The tail of a comet formed of dust blown outward by the pressure of sunlight. (See *gas tail*.)

dwarf planet An object that orbits the Sun and has pulled itself into a spherical shape but has not cleared its orbital lane of other objects. Pluto is a dwarf planet.

dynamo effect The process by which a rotating, convecting body of conducting matter, such as Earth's core, can generate a magnetic field.

east point The point on the eastern horizon exactly halfway between the north point and the south point; exactly east.

eccentric (noun) An off-center circular path, such as the orbits of other celestial bodies around Earth in the Ptolemaic model. (Note that the adjective "eccentric" refers instead to an ellipse that is not a perfect circle.)

eccentricity *(e)* A measure of the flattening of an ellipse. An ellipse of $e = 0$ is circular. The closer to 1 that e becomes, the more flattened the ellipse.

eclipse season That period when the Sun is near a node of the Moon's orbit and eclipses are possible.

eclipse year The time the Sun takes to circle the sky and return to a node of the Moon's orbit; 346.62 days.

eclipsing binary system A binary star system in which the stars eclipse each other.

ecliptic The apparent path of the Sun around the sky.

ejecta Pulverized rock scattered by meteorite impacts on a planetary surface.

electromagnetic radiation Changing electric and magnetic fields that travel through space and transfer energy from one place to another—for example, light, radio waves, and the like.

electron Low-mass atomic particle carrying a negative charge.

ellipse A closed curve enclosing two points (foci) such that the total distance from one focus to any point on the curve back to the other focus equals a constant. The orbits of the planets in our Solar System are ellipses.

elliptical galaxy A galaxy that is round or elliptical in outline; it contains little gas and dust, no disk or spiral arms, and few hot, bright stars.

emission spectrum A spectrum containing emission lines; also called a bright-line spectrum.

emission line A bright line in a spectrum caused by the emission of photons from atoms.

emission nebula A cloud of glowing gas excited by ultraviolet radiation from hot stars.

empirical Description of a phenomenon without explaining why it occurs. Kepler's laws of planetary motion are empirical laws.

energy The capacity of a natural system to perform work—for example, thermal energy.

energy level One of a number of states an electron may occupy in an atom, depending on its binding energy.

energy transport The law of energy transport states that energy must flow from hot regions to cool regions by conduction, convection, or radiation.

enzyme Special protein that controls processes in an organism.

epicycle The small circle followed by a planet in the Ptolemaic theory. The center of the epicycle follows a larger circle (deferent) around Earth.

equant In the Ptolemaic theory, the point off-center in the deferent from which the center of the epicycle appears to move uniformly.

equatorial mount A telescope mounting that allows motion parallel to and perpendicular to the celestial equator.

escape velocity (V_e) The initial velocity an object needs to escape from the surface of a celestial body.

evening star Any planet visible in the sky just after sunset.

event horizon The boundary of the region of a black hole from which no radiation may escape. No event that occurs within the event horizon is visible to a distant observer.

evolutionary hypothesis Explanation for natural events that involves gradual changes as opposed to sudden catastrophic changes—for example, the formation of the planets in the gas cloud around the forming Sun.

excited atom An atom in which an electron has moved from a lower to a higher orbit.

expanding universe The idea, supported by observed redshifts of galaxies, that space is stretching, carrying galaxies and galaxy clusters away from each other.

extrasolar planet A planet orbiting a star other than the Sun.

extremophile An organism that can survive in an extreme environment, for example, very low or high temperatures, high acidity, and extreme dryness.

eyepiece A short-focal-length lens used to enlarge the image in a telescope; the lens nearest the eye.

fall A meteorite seen to fall. (See *find*.)

false-color image A representation of graphical data in which the colors are altered or added to reveal details.

field A way of explaining action at a distance; a particle produces a field of influence (gravitational, electric, or magnetic) to which another particle in the field responds.

field of view The area visible in an image; usually given as the diameter of the region.

filament (1) On the Sun, a prominence seen silhouetted against the solar surface. (2) A linear region containing many galaxies and galaxy clusters, part of the large-scale structure of the Universe.

filtergram An image (usually of the Sun) taken in the light of a specific region of the spectrum—for example, an H-alpha filtergram.

find A meteorite that is found but was not seen to fall. (See *fall*.)

first principle A first principle is an idea considered so obviously true that the idea does not need to be questioned. Classical philosophers accepted as a first principle that Earth was the unmoving center of the Universe.

flare A violent eruption on the Sun's surface.

flat universe A model universe in which space-time is not curved.

flatness problem In cosmology, the circumstance that the early Universe must have contained almost exactly the right amount of matter to close space-time (to make space-time flat).

flocculent galaxy A galaxy that has a woolly appearance.

flux A measure of the flow of energy onto or through a surface. Usually applied to light.

focal length The distance from a lens to the point where it focuses parallel rays of light.

folded mountain range A long range of mountains formed by the compression of a planet's crust—for example, the Andes on Earth.

forbidden line A spectral line that does not occur in the laboratory because it depends on an atomic transition that is highly unlikely.

forward scattering The optical property of finely divided particles to preferentially direct light in the original direction of the light's travel.

free-fall collapse The early contraction of a gas cloud to form a star during which internal pressure is too low to resist contraction.

frequency (n) The number of times a given event occurs in a given time; for a wave, the number of cycles that pass the observer in 1 second.

frost line In the solar nebula, the boundary beyond which water vapor and other compounds could form ice particles.

FU Orionis phenomenon When a pre-main-sequence star displays extreme short-term changes in brightness and spectral type.

galactic cannibalism The theory that large galaxies absorb smaller galaxies.

galactic corona The low-density extension of the halo of a galaxy; now suspected to extend many times the visible diameter of the galaxy.

galactic fountain A region of the galaxy's disk in which gas heated by supernova explosions throws gas out of the disk where it can fall back and spread metals through the disk.

galaxy A very large collection of gas, dust, and stars orbiting a common center of mass. The Sun and Earth are located in the Milky Way Galaxy.

Galilean moons The four largest satellites of Jupiter, named after their discoverer, Galileo.

gamma-ray Electromagnetic wave with extremely short wavelength, high frequency, and large photon energy.

gamma-ray burst (GRB) A sudden burst of gamma-rays thought to be associated with neutron stars and black holes.

gas tail The tail of a comet produced by gas blown outward by the solar wind. (See *dust tail*.)

gene A unit of DNA containing genetic information that influences a particular inherited trait.

general theory of relativity Einstein's more sophisticated theory of space and time, which describes gravity as a curvature of space-time.

geocentric universe A model universe with Earth at the center, such as the Ptolemaic universe.

geosynchronous satellite An Earth satellite in an eastward orbit whose period is 24 hours. A satellite in such an orbit remains above the same spot on Earth's surface.

giant Large, cool, highly luminous star in the upper right of the H–R diagram, typically 10 to 100 times the diameter of the Sun.

giant molecular cloud Very large, cool cloud of dense gas in which stars form.

global warming The gradual increase in the surface temperature of Earth caused by human modifications to Earth's atmosphere.

globular cluster A star cluster containing 50,000 to 1 million stars in a sphere about 75 ly in diameter; generally old, metal poor, and found in the spherical component of the galaxy.

gossamer ring The dimmest part of Jupiter's ring, produced by dust particles orbiting near small moons.

grand design galaxy Galaxy with a high-contrast, simple, two-arm spiral pattern.

grand unified theory (GUT) Theory that attempts to unify (describe in a similar way) the electromagnetic, weak, and strong forces of nature.

granulation The fine structure visible on the solar surface caused by rising currents of hot gas and sinking currents of cool gas below the surface.

grating A piece of material in which numerous microscopic parallel lines are scribed; light encountering a grating is dispersed to form a spectrum.

gravitational collapse The stage in the formation of a massive planet when it grows massive enough to begin capturing gas directly from the nebula around it.

gravitational lensing The effect of the focusing of light from a distant galaxy or quasar by an intervening galaxy to produce multiple images of the distant body.

gravitational radiation As predicted by general relativity, expanding waves in a gravitational field that transport energy through space.

gravitational redshift The lengthening of the wavelength of a photon its escape from a gravitational field.

greenhouse effect The process by which a carbon dioxide atmosphere traps heat and raises the temperature of a planetary surface.

grooved terrain Region of the surface of Ganymede consisting of bright, parallel grooves.

ground state The lowest permitted electron orbit in an atom.

H I cloud An interstellar cloud of neutral hydrogen.

habitable zone The region around a star within which an orbiting planet can have surface temperatures allowing liquid water.

half-life The time required for half of the atoms in a radioactive sample to decay.

halo The spherical region of a spiral galaxy containing a thin scattering of stars, star clusters, and small amounts of gas.

heat Energy flowing from a warm body to a cool body by the agitation of particles such as atoms or molecules.

heat of formation In planetology, the heat released by the infall of matter during the formation of a planetary body.

heavy bombardment The period of intense meteorite impacts early in the formation of the planets, when the Solar System was filled with debris.

heliocentric universe A model universe with the Sun at the center, such as the Copernican universe.

heliopause The surface at which the solar wind collides with the interstellar medium. Effectively, the outer boundary of the Solar System.

helioseismology The study of the interior of the Sun by the analysis of its modes of vibration.

helium flash The explosive ignition of helium burning that takes place in some giant stars.

Herbig–Haro object A small nebula associated with star formation that varies irregularly in brightness.

Hertzsprung–Russell (H–R) diagram A plot of the intrinsic brightness versus the surface temperature of stars; it separates the effects of temperature and surface area on stellar luminosity; commonly absolute magnitude versus spectral type but also luminosity versus surface temperature or color.

Hesperian period On Mars, the geological era from the decline of heavy cratering and lava flows and the melting of subsurface ice to form the outflow channels.

H II region A region of ionized hydrogen around a hot star.

Hirayama family Family of asteroids with or bits of similar size, shape, and orientation; thought to be fragments of larger bodies.

homogeneous The property of being uniform. In cosmology, the characteristic of the Universe in which, on the large scale, matter is uniformly spread through the Universe.

horizon The line that marks the apparent intersection of Earth and the sky.

horizon problem In cosmology, the circumstance that the primordial background radiation seems much more isotropic than could be explained by the standard Big Bang theory.

horizontal branch In the H–R diagram of a globular cluster, the sequence of stars extending from the red giants toward the blue side of the diagram; includes RR Lyrae stars.

hot dark matter Invisible matter in the Universe composed of low-mass, high-velocity particles such as neutrinos.

hot Jupiter A massive and presumably Jovian planet that orbits close to its star and consequently has a high temperature.

hot spot In radio astronomy, a bright spot in a radio lobe.

H–R diagram See *Hertzsprung–Russell diagram*.

Hubble constant (*H*) A measure of the rate of expansion of the Universe; the average value of velocity of recession divided by distance; about 70 km/s/megaparsec.

Hubble law The linear relation between the distance to a galaxy and its radial velocity.

Hubble time An upper limit on the age of the Universe derived from the Hubble constant.

hydrostatic equilibrium The balance between the weight of the material pressing downward on a layer in a star and the pressure in that layer.

hypernova The explosion produced as a very massive star collapses into a black hole; thought to be responsible for at least some gamma-ray bursts.

hypothesis A conjecture, subject to further tests, that accounts for a set of facts.

inertia The property of matter that resists changes in motion.

inflationary universe A version of the Big Bang theory that includes a rapid expansion when the Universe was very young.

infrared (IR) radiation Electromagnetic radiation with wavelengths intermediate between visible light and radio waves.

inner Lagrange (L_1) point The point of gravitational equilibrium between two orbiting stars through which matter can flow from one star to the other.

instability strip The region of the H–R diagram in which stars are unstable to pulsation; a star passing through this strip becomes a variable star.

intercloud medium The hot, low-density gas between cooler clouds in the interstellar medium.

intercrater plain The relatively smooth terrain on Mercury.

interferometry The observing technique in which separated telescopes combine to produce a virtual telescope with the resolution of a much-larger-diameter telescope.

intergalactic medium Material (gas and dust) between galaxies; analogous to the interstellar medium.

International Astronomical Union (IAU) An international society of astronomers that, among other activities, decides definitions and naming conventions for celestial objects and surface features. The IAU defined the constellation boundaries in 1930 and reclassified Pluto as a dwarf planet in 2006.

interstellar absorption line One of the dark lines in some stellar spectra that are formed by interstellar gas.

interstellar dust Microscopic solid grains in the interstellar medium.

interstellar extinction The dimming of starlight by gas and dust in the interstellar medium.

interstellar medium (ISM) The gas and dust distributed between the stars.

interstellar reddening The process in which dust scatters blue light out of starlight and makes the stars look redder.

intrinsic brightness The true brightness of an object independent of its distance. Also referred to as *luminosity*.

intrinsic variable star A variable star driven to pulsate by processes in its interior.

inverse square law The rule that the strength of an effect (such as gravity) decreases in proportion as the distance squared increases.

Io flux tube A tube of magnetic lines and electric currents connecting Io and Jupiter.

Io plasma torus The doughnut-shaped cloud of ionized gas that encloses the orbit of Jupiter's moon Io.

ion An atom that has lost or gained one or more electrons.

ionization The process in which atoms lose or gain electrons.

iron meteorite A meteorite composed mainly of iron–nickel alloy.

irregular galaxy A galaxy with a chaotic appearance, large clouds of gas and dust, and both population I and population II stars, but without spiral arms.

irregular satellite A moon with an orbit that has large eccentricity or high inclination to the equator of its parent planet or is retrograde. Irregular moons are thought to have been captured.

isotopes Atoms that have the same number of protons but a different number of neutrons.

isotropic The condition of being uniform in all directions. In cosmology, the characteristic of the Universe by which, in its general properties, it looks the same in every direction.

joule (J) A unit of energy equivalent to a force of 1 newton acting over a distance of 1 meter; 1 joule per second equals 1 watt of power.

Jovian planet Jupiter-like planet with large diameter and low density.

Jovian problem The puzzle that protoplanetary disks around young stars don't seem to survive long enough to form Jovian planets by condensation, accretion, and gravitational collapse, yet Jovian-mass extrasolar planets are common. See also *extrasolar planet*.

jumbled terrain Disturbed regions of the Moon's surface opposite the locations of the Imbrium Basin and Mare Orientale, possibly focusing of seismic waves from the large impacts that formed those basins.

Kelvin temperature scale The temperature, in Celsius (centigrade) degrees, measured above absolute zero.

Keplerian motion Orbital motion in accord with Kepler's laws of planetary motion.

kiloparsec (kpc) A unit of distance equal to 1000 pc, or 3260 ly.

Kirchhoff's laws A set of laws that describe the absorption and emission of light by matter.

Kirkwood gaps Regions in the asteroid belt in which there are very few asteroids; caused by orbital resonances with Jupiter.

Kuiper Belt The collection of icy planetesimals that orbit in a region from just beyond Neptune out to about 50.

Kuiper Belt object (KBO) An object in the Kuiper Belt, a region beyond Neptune's orbit containing planetesimals remaining from the formation of the Solar System. Pluto is one of the largest Kuiper Belt objects.

L dwarf A type of star that is even cooler than the M stars.

Lagrange point Point of stability in the orbital plane of a binary star system, planet, or moon. One is located 60° ahead and one 60° behind the orbiting bodies; another is located between the orbiting bodies.

large-impact hypothesis The hypothesis that the Moon formed from debris ejected during a collision between Earth and a large planetesimal.

large-scale structure The distribution of galaxy clusters and superclusters in walls and filaments surrounding voids mostly empty of galaxies.

laser guide star An artificial star image produced by a laser pointing up into Earth's atmosphere, used as a reference in adaptive optics systems. See also *adaptive optics*.

late heavy bombardment The surge in cratering impacts in the Solar System that occurred about 3.8 billion years ago.

law of energy conservation One of the basic laws of stellar structure. The amount of energy flowing out of the top of a shell must equal the amount coming in at the bottom plus whatever energy is generated within the shell.

law of mass conservation One of the basic laws of stellar structure. The total mass of the star must equal the sum of the masses of the shells, and the mass must be distributed smoothly throughout the star.

lenticular galaxy A galaxy with an obvious disk and nuclear bulge but little or no visible gas and dust and no hot, bright stars.

light curve A graph of brightness versus time commonly used in analyzing variable stars and eclipsing binaries.

light pollution The illumination of the night sky by waste light from cities and outdoor lighting, which prevents the observation of faint objects.

light-gathering power The ability of a telescope to collect light; proportional to the area of the telescope's objective lens or mirror.

lighthouse model The explanation of a pulsar as a spinning neutron star sweeping beams of radio radiation around the sky.

light-year (ly) The distance light travels in one year.

limb The edge of the apparent disk of a body, as in "the limb of the Moon."

limb darkening The decrease in brightness of the Sun or other body from its center to its limb.

line of nodes The line across an orbit connecting the nodes; commonly applied to the orbit of the Moon.

liquid metallic hydrogen A form of hydrogen under high pressure that is a good electrical conductor.

lobate scarp A curved cliff such as those found on Mercury.

local bubble A region of high-temperature, low-density gas in the interstellar medium in which the Sun happens to be located.

Local Group The cluster of about 40 nearby galaxies of which our Milky Way Galaxy is a member.

look-back time The amount by which you look into the past when you look at a distant galaxy; a time equal to the distance to the galaxy in light-years.

luminosity (*L*) The total amount of energy a star radiates in 1 second.

luminosity class A category of stars of similar luminosity; determined by the widths of lines in their spectra.

lunar eclipse The darkening of the Moon when it moves through Earth's shadow.

lunar phase The appearance of the Moon from Earth in terms of which portion is lit by the Sun versus which portion is dark, which changes in a regular monthly cycle.

Lyman series Spectral lines in the ultraviolet spectrum of hydrogen produced by transitions whose lowest orbit is the ground state.

magnetar A class of neutron stars that have exceedingly strong magnetic fields; thought to be responsible for soft gamma-ray repeaters.

magnetic carpet The widely distributed, low-level magnetic field extending up through the Sun's visible surface.

magnetosphere The volume of space around a planet within which the motion of charged particles is dominated by the planetary magnetic field rather than the solar wind.

magnifying power The ability of a telescope to make an image larger.

magnitude scale The astronomical brightness scale; the larger the number, the fainter the star.

magnitude–distance formula The mathematical formula that relates the apparent magnitude and absolute magnitude of a star to its distance.

main sequence The region of the H–R diagram running from upper left to lower right, which includes roughly 90 percent of all stars.

mantle The layer of dense rock and metal oxides that lies between the molten core and Earth's surface; also, similar layers in other planets.

mare (plural: *maria*) One of the lunar lowlands filled by successive flows of dark lava; from the Latin word for "sea."

mass A measure of the amount of matter making up an object.

mass–luminosity relation The more massive a star is, the more luminous it is.

Maunder butterfly diagram A graph showing the latitude of sunspots versus time; first plotted by W. W. Maunder in 1904.

Maunder minimum A period of less numerous sunspots and other solar activity from 1645 to 1715.

megaparsec (Mpc) A unit of distance equal to 1 million pc.

meridional flow Flows of material from the Sun's equator toward the poles. These flows carry portions of the solar magnetic field and are understood to be one mechanism that helps drive the 11- and 22-year solar activity cycles.

metal In astronomical usage, any atom heavier than helium.

metastable level An atomic energy level from which an electron takes a long time to decay; responsible for producing forbidden lines.

meteor A small bit of matter heated by friction to incandescent vapor as it falls into Earth's atmosphere.

meteor shower An event lasting for hours or days in which the number of meteors entering Earth's atmosphere suddenly increases. The meteors in a shower have a common origin and are traveling through space on nearly parallel paths.

meteorite A meteor that has survived its passage through the atmosphere and strikes the ground.

meteoroid A meteor in space before it enters Earth's atmosphere.

microlensing Brightening of a background star due to focusing of its light by the gravity of a foreground extrasolar planet, allowing the planet to be detected and some of its characteristics measured.

micrometeorite Meteorite of microscopic size.

microwave Electromagnetic wave with wavelength, frequency, and photon energy intermediate between infrared and radio waves.

midocean rise One of the undersea mountain ranges that push up from the seafloor in the center of the oceans.

Milankovitch hypothesis The hypothesis that small changes in Earth's orbital and rotational motions cause the ice ages.

Milky Way Galaxy The spiral galaxy containing the Sun; visible at night as the hazy band of light that circles the sky, produced by the combined light of billions of stars in our galaxy.

Miller experiment An experiment that reproduced the conditions under which life began on Earth and amino acids and other organic compounds were manufactured.

millisecond pulsar A pulsar with a period of approximately a millisecond, a thousandth of a second.

molecular cloud An interstellar gas cloud that is dense enough for the formation of molecules; discovered and studied through the radio emissions of such molecules.

molecule Two or more atoms bonded together.

momentum The tendency of a moving object to continue moving; mathematically, the product of mass and velocity.

morning star Any planet visible in the sky just before sunrise.

M-type asteroid A type of asteroid with relatively high reflectivity and grayish color, probably composed primarily of metal.

multicellular Describes an organism composed of many cells.

multiringed basin Very large impact basin in which there are concentric rings of mountains.

mutation Offspring born with altered DNA.

nadir The point on the bottom of the celestial sphere directly under your feet.

nanometer (nm) A unit of length equal to 10^{-9} m.

natural law A conjecture about how nature works in which scientists have overwhelming confidence.

natural motion In Aristotelian physics, the motion of objects toward their natural places—fire and air upward and earth and water downward.

natural selection The process by which the best traits are passed on, allowing the most able to survive.

neap tide Ocean tide of low amplitude occurring at first- and third-quarter moon.

Near-Earth Object (NEO) An asteroid or comet in an orbit that passes near or intersects Earth's orbit that could potentially collide with Earth.

nebula A cloud of gas and dust in space.

nebular hypothesis The proposal that the Solar System formed from a rotating cloud of gas.

neutrino A neutral, massless atomic particle that travels at or nearly at the speed of light.

neutron An atomic particle with no charge and about the same mass as a proton.

neutron star A small, highly dense star composed almost entirely of tightly packed neutrons; radius about 10 km.

Newtonian focus The focal arrangement of a reflecting telescope in which a diagonal mirror reflects light out the side of the telescope tube for easier access.

Noachian period On Mars, the era that extends from the formation of the crust to the end of heavy cratering and includes the formation of the valley networks.

node A point where an object's orbit passes through the plane of Earth's orbit.

nonbaryonic matter In cosmology, a suspected component of the dark matter composed of matter that does not contain protons and neutrons.

north celestial pole The point on the celestial sphere directly above Earth's North Pole.

north point The point on the horizon directly below the north celestial pole; exactly north.

nova From the Latin "new," a sudden brightening of a star, making it appear as a "new" star in the sky; thought to be associated with eruptions on white dwarfs in binary systems.

nuclear fission Reaction that splits nuclei into less massive fragments.

nuclear fusion Reaction that joins the nuclei of atoms to form more massive nuclei.

nucleosynthesis The production of elements heavier than helium by the fusion of atomic nuclei in stars and during supernovae explosions.

nucleus (of a comet) A lump of fragile, icy material that makes up the solid portion of a comet.

nucleus (of an atom) The central core of an atom containing protons and neutrons; carries a net positive charge.

OB association A loosely bound cluster of young stars having spectral types O and B, indicating a region of relatively recent star formation.

oblateness The flattening of a spherical body, usually caused by rotation.

observable universe The part of the Universe that is visible from Earth's location in space and time.

occultation The passage of a larger body in front of a smaller body.

Olbers's paradox The conflict between observation and theory as to why the night sky should or should not be dark.

Oort Cloud The cloud of icy bodies—extending from the outer part of our Solar System out to roughly

100,000 from the Sun—that acts as the source of most comets.

opacity The resistance of a gas to the passage of radiation.

open cluster A cluster of 10 to 10,000 stars with an open, transparent appearance and stars not tightly grouped; usually relatively young and located in the disk of the galaxy.

open orbit An orbit that does not return to its starting point; an escape orbit. (See *closed orbit*.)

open universe A model universe in which the average density is less than the critical density needed to halt the expansion.

optical telescope A telescope that gathers and focuses visible light. See also *radio telescope*.

outflow channel Geological feature on Mars that appears to have been caused by sudden flooding.

outgassing The release of gases from a planet's interior.

ovoid Geological feature on Uranus's moon Miranda thought to be produced by circulation in the solid icy mantle and crust.

ozone layer In Earth's atmosphere, a layer of oxygen ions (O_3) lying 15 to 30 km high that protects the surface by absorbing ultraviolet radiation.

paradigm A commonly accepted set of scientific ideas and assumptions.

parallax The apparent change in the position of an object as a result of a change in the location of the observer. Astronomical parallax is measured in seconds of arc.

parsec (pc) The distance to a hypothetical star whose parallax is 1 second of arc; 1 pc = 206,265 AU = 3.26 ly.

partial lunar eclipse A lunar eclipse in which the Moon does not completely enter Earth's shadow.

partial solar eclipse A solar eclipse in which the Moon does not completely cover the Sun.

Paschen series Spectral lines in the infrared spectrum of hydrogen produced by transitions whose lowest orbit is the third.

passing star hypothesis The proposal that our Solar System formed when two stars passed near each other and material was pulled out of one to form the planets.

path of totality The track of the Moon's umbral shadow over Earth's surface during an eclipse. The Sun is totally eclipsed as seen from within this path.

penumbra The portion of a shadow that is only partially shaded, typically used to describe the shadow of Earth or Moon during an eclipse.

penumbral lunar eclipse A lunar eclipse in which the Moon enters the penumbra of Earth's shadow but does not reach the umbra.

perigee The orbital point of closest approach to Earth.

perihelion The orbital point of closest approach to the Sun.

period–luminosity relation The relation between period of pulsation and intrinsic brightness among Cepheid variable stars.

permitted orbit One of the energy levels in an atom that an electron may occupy.

photographic plate An old-fashioned means of recording astronomical images and photometric information on a photographic emulsion coating a glass plate. See also *array detector* and *charge-coupled device*.

photometer An instrument attached to a telescope for the purpose of precisely measuring the brightness of stars or other objects at one or more wavelengths.

photon A quantum of electromagnetic energy; carries an amount of energy that depends inversely on its wavelength.

photosphere The bright visible surface of the Sun.

planet A nonluminous object, larger than a comet or asteroid, that orbits a star.

planetary nebula An expanding shell of gas ejected from a star during the latter stages of its evolution.

planetesimal One of the small bodies that formed from the solar nebula and eventually grew into protoplanets.

plastic A material with the properties of a solid but capable of flowing under pressure.

plate tectonics The constant destruction and renewal of Earth's surface by the motion of sections of crust.

plutino One of the icy Kuiper Belt objects that, like Pluto, are caught in a 3:2 orbital resonance with Neptune.

polar axis In an equatorial telescope mounting, the axis that is parallel to Earth's axis of rotation.

polarity Orientation and strength of a magnetic field's manifestation as north and south poles. Also applies to an electrical field's manifestation as positive and negative charges.

poor cluster An irregularly shaped cluster that contains fewer than 1000 galaxies, many spiral, and no giant ellipticals.

Population I star Star rich in atoms heavier than helium; nearly always a relatively young star found in the disk of a galaxy.

Population II star Star poor in atoms heavier than helium; nearly always a relatively old star found in the halo, globular clusters, or the nuclear bulge of a galaxy.

positron The antiparticle of the electron.

precession The slow change in the direction of Earth's axis of rotation; one cycle takes nearly 26,000 years.

pressure A force exerted over a surface; expressed as force per unit area.

pressure *(P)* wave In geophysics, a mechanical wave of compression and rarefaction that travels through Earth's interior.

pressure–temperature thermostat The dependence of gas pressure on gas temperature, resulting in stability and regulation of energy production in the cores of normal stars.

primary lens The main lens in a refracting telescope.

primary mirror The main optical element in an astronomical telescope. The large lens at the top of the telescope tube or the large mirror at the bottom.

prime focus The point at which the objective mirror forms an image in a reflecting telescope.

primordial soup The rich solution of organic molecules in Earth's first oceans.

prograde Rotation or revolution in the direction in common with most such motions in the Solar System. See also *retrograde*.

prominence Eruption on the solar surface; visible during total solar eclipses.

proper motion The rate at which a star moves across the sky; measured in seconds of arc per year.

protein Complex molecule composed of amino acid units.

protogalaxy A cloud of gas collapsing gravitationally to become a galaxy. Evidently, protogalaxies only existed in the early history of the Universe.

proton A positively charged atomic particle contained in the nucleus of an atom; the nucleus of a hydrogen atom.

proton–proton chain A series of three nuclear reactions that build a helium atom by adding together protons; the main energy source in the Sun.

protoplanet Massive object resulting from the coalescence of planetesimals in the solar nebula and destined to become a planet.

protostar A collapsing cloud of gas and dust destined to become a star.

protostellar disk A gas cloud around a forming star flattened by its rotation.

pseudoscience A subject that claims to obey the rules of scientific reasoning but does not. Examples include astrology, crystal power, and pyramid power.

pulsar A source of short, precisely timed radio bursts; thought to be a spinning neutron star.

pulsar wind The flow of high-energy particles that carries most of the energy away from a spinning neutron star.

quantum mechanics The study of the behavior of atoms and atomic particles.

quasar (quasi-stellar object, or QSO) Small, powerful source of energy thought to be the active core of a very distant galaxy.

quintessence The proposed energy of empty space that causes the acceleration of the expanding Universe.

radial velocity (V_r) That component of an object's velocity directed away from or toward Earth.

radiant The point in the sky from which meteors in a shower seem to come.

radiation pressure The force exerted on the surface of a body by its absorption of light. Small particles floating in the Solar System can be blown outward by the pressure of the sunlight.

radiative zone The region inside a star where energy is carried outward as photons.

radio galaxy A galaxy that is a strong source of radio signals.

radio telescope A telescope that gathers and focuses electromagnetic energy with microwave and radio wavelengths. See also *optical telescope*.

radio wave Electromagnetic wave with extremely long wavelength, low frequency, and small photon energy.

ray Ejecta from a meteorite impact, forming white streamers radiating from some lunar craters.

recombination The stage within a million years of the Big Bang when the gas became transparent to radiation.

reconnection event The process in the Sun's atmosphere by which opposing magnetic fields combine and release energy to power solar flares.

red dwarf Cool, low-mass star on the lower main sequence.

reddening curve A graph that displays how much starlight is blocked by interstellar dust as a function of wavelength. The average size of interstellar grains can be inferred from these data.

redshift The lengthening of the wavelengths of light seen when the source and observer are receding from each other.

reflecting telescope A telescope that uses a concave mirror to focus light into an image.

reflection nebula A nebula produced by starlight reflecting off dust particles in the interstellar medium.

refracting telescope A telescope that forms images by bending (refracting) light with a lens.

regolith A soil made up of crushed rock fragments.

regular satellite A moon with an orbit that has small eccentricity, low inclination to the equator of its parent planet, and is prograde. Regular moons are thought to have formed with their respective planets rather than having been captured.

re-ionization The stage in the early history of the Universe when ultraviolet photons from the first stars ionized the gas filling space.

relative age The age of a geological feature as referred to other features. For example, relative ages reveal that the lunar maria are younger than the highlands.

resolving power The ability of a telescope to reveal fine detail; depends on the diameter of the telescope objective.

resonance The coincidental agreement between two periodic phenomena; commonly applied to agreements between orbital periods, which can make orbits more stable or less stable.

retrograde motion (apparent) The apparent backward (westward) motion of planets as seen against the background of stars.

retrograde motion (absolute) Motion in the opposite direction from other bodies. For example, Venus rotates in the opposite direction from other planets, and some moons orbit in the opposite direction to the spin of their planets.

revolution The motion of an object in a closed path about a point outside its volume; Earth revolves around the Sun.

rich cluster A cluster containing more than 1000 galaxies, mostly elliptical, scattered over a volume about 3 in diameter.

rift valley A long, straight, deep valley produced by the separation of crustal plates.

ring galaxy A galaxy that resembles a ring around a bright nucleus; thought to be the result of a head-on collision of two galaxies.

RNA (ribonucleic acid) A long carbon-chain molecule that uses the information stored in DNA to manufacture complex molecules necessary to the organism.

Roche limit The minimum distance between a planet and a satellite that holds itself together by its own gravity. If a satellite's orbit brings it within its planet's Roche limit, tidal forces will pull the satellite apart.

Roche lobe In a system with two bodies orbiting each other, the volume of space dominated by the gravitation of one of the bodies.

Roche surface In a system with two bodies orbiting each other, the outer boundary of the volume of space dominated by the gravitation of one of the bodies.

rotation The turning of a body about an axis that passes through its volume; Earth rotates on its axis.

rotation curve A graph of orbital velocity versus radius in the disk of a galaxy.

rotation curve method The procedure for finding the mass of a galaxy from its rotation curve.

RR Lyrae variable star Variable star with a period of 12 to 24 hours; common in some globular clusters.

runaway greenhouse effect A "vicious cycle" (positive feedback) in which a planet's greenhouse effect causes enhanced release or retention of greenhouse gases that increases the greenhouse effect, and so on. There is evidence that Venus suffered a runaway greenhouse effect early in its history with the result that its oceans boiled away and were destroyed by solar.

Saros cycle An 18-year 11 ⅓-day period after which the pattern of lunar and solar eclipses repeats.

Schmidt-Cassegrain focus The optical design of a reflecting telescope in which a thin correcting lens is placed at the top of a Cassegrain telescope.

Schwarzschild radius (R_S) The radius of the event horizon around a black hole.

scientific argument An honest, logical discussion of observations and theories intended to reach a valid conclusion.

scientific method The reasoning style by which scientists test theories against evidence to understand how nature works.

scientific model An intellectual concept designed to help you think about a natural process without necessarily being a conjecture of truth.

scientific notation The system of recording very large or very small numbers by using powers of 10.

Search for Extra Terrestrial Intelligence (SETI) The field of study devoted to searching for extraterrestrial radio signals as evidence of alien intelligence.

secondary crater A crater formed by the impact of debris ejected from a larger crater.

secondary mirror In a reflecting telescope, the mirror that reflects the light to a point of easy observation.

seeing Atmospheric conditions on a given night. When the atmosphere is unsteady, producing blurred images, the seeing is said to be poor.

seismic wave A mechanical vibration that travels through Earth; usually caused by an earthquake.

seismograph An instrument that records seismic waves.

selection effect An influence on the probability that certain phenomena will be detected or selected, which can alter the outcome of a survey.

self-sustaining star formation The process by which the birth of stars compresses the surrounding gas clouds and triggers the formation of more stars; proposed to explain spiral arms.

semimajor axis (a) Half of the longest axis of an ellipse.

Seyfert galaxy An otherwise normal spiral galaxy with an unusually bright, small core that fluctuates in brightness; thought to indicate the core is erupting.

Shapley–Curtis Debate The 1920 debate between Harlow Shapley and Heber Curtis over the nature of the spiral nebulae.

shear (S) wave A mechanical wave that travels through Earth's interior by the vibration of particles perpendicular to the direction of wave travel.

shepherd satellite A satellite that, by its gravitational field, confines particles to a planetary ring.

shield volcano Wide, low-profile volcanic cone produced by highly liquid lava.

shock wave A sudden change in pressure that travels as an intense sound wave.

sidereal drive The motor and gears on a telescope that turn it westward to keep it pointed at a star.

sidereal period The period of rotation or revolution of an astronomical body relative to the stars.

singularity The object of zero radius into which the matter in a black hole is thought to fall.

sinuous rille A narrow, winding valley on the Moon caused by ancient lava flows along narrow channels.

small-angle formula The mathematical formula that relates an object's linear diameter and distance to its angular diameter.

smooth plain Apparently young plain on Mercury formed by lava flows at or soon after the formation of the Caloris Basin.

solar constant A measure of the energy output of the Sun; the total solar energy striking 1 m^2 just above Earth's atmosphere in 1 second.

solar eclipse The event that occurs when the Moon passes directly between Earth and the Sun, blocking your view of the Sun.

solar nebula theory The proposal that the planets formed from the same cloud of gas and dust that formed the Sun.

Solar System The Sun and the nonluminous objects that orbit it, including the planets, comets, and asteroids.

solar wind Rapidly moving atoms and ions that escape from the solar corona and blow outward through the Solar System.

south celestial pole The point of the celestial sphere directly above Earth's South Pole.

south point The point on the horizon directly above the south celestial pole; exactly south.

special theory of relativity The first of Einstein's theories of relativity, which deals with uniform motion.

spectral line A dark or bright line that crosses a spectrum at a specific wavelength.

spectral sequence The arrangement of spectral classes (O, B, A, F, G, K, M) ranging from hot to cool.

spectral type or class A star's position in the temperature classification system O, B, A F, G, K, and M. Based on the appearance of the star's spectrum.

spectrograph A device that separates light by wavelength to produce a spectrum.

spectroscopic binary system A star system in which the stars are too close together to be visible separately. You see a single point of light, and only by taking a spectrum can you determine that there are two stars.

spectroscopic parallax The method of determining a star's distance by comparing its apparent magnitude

with its absolute magnitude, as estimated from its spectrum.

spectrum An arrangement of electromagnetic radiation in order of wavelength or frequency.

spherical component The part of the galaxy including all matter in a spherical distribution around the center (the halo and nuclear bulge).

spicule Small, flamelike projection in the chromosphere of the Sun.

spiral arm Long, spiral pattern of bright stars, star clusters, gas, and dust that extends from the center to the edge of the disk of spiral galaxies.

spiral density wave theory The conjecture that spiral arms in disk galaxies are caused by a pressure wave that rotates slowly around the galaxy, triggering star formation by compressing interstellar gas clouds.

spiral galaxy A galaxy with an obvious disk component containing gas; dust; hot, bright stars; and spiral arms.

spiral nebula Nebulous object with a spiral appearance observed in early telescopes; later recognized as a spiral galaxy.

spiral tracer Object used to map the spiral arms, for example O and B associations, open clusters, clouds of ionized hydrogen, and some types of variable stars.

sporadic meteor A meteor not part of a meteor shower.

spring tide Ocean tide of high amplitude that occurs at full and new moon.

star A celestial object composed of gas held together by its own gravity and supported by nuclear fusion occurring in its interior.

starburst galaxy A bright blue galaxy in which many new stars are forming, thought to be caused by collisions between galaxies.

star-formation pillar The column of gas produced when a dense core of gas protects the nebula behind it from the energy of a nearby hot star that is evaporating and driving away a star-forming nebula.

Stefan-Boltzmann law The mathematical relation between the temperature of a blackbody (an ideal radiator) and the amount of energy emitted per second from 1 square meter of its surface.

stellar model A table of numbers representing the conditions in various layers within a star.

stellar parallax (p) A measure of stellar distance. (See *parallax*.)

stellar wind Hot gases blowing outward from the surface of a star. The equivalent for another star of the solar wind.

stony-iron meteorite A meteorite that is a mixture of stone and iron.

stony meteorite A meteorite composed of silicate (rocky) material.

stromatolite A layered fossil formation caused by ancient mats of algae or bacteria that build up mineral deposits season after season.

strong nuclear force One of the four forces of nature; the strong force binds protons and neutrons together in atomic nuclei.

S-type asteroid A type of asteroid common in the inner asteroid belt, with relatively high reflectivity and reddish color, probably composed of rocky material.

subduction zone A region of a planetary crust where a tectonic plate slides downward.

subsolar point The point on a planet that is directly below the Sun.

summer solstice The point on the celestial sphere where the Sun is at its most northerly point; also, the time when the Sun passes this point, about June 22, and summer begins in the Northern Hemisphere.

sunspot Relatively dark spot on the Sun that contains intense magnetic fields.

supercluster A cluster of galaxy clusters.

supergiant Exceptionally luminous star, 10 to 1000 times the Sun's diameter.

supergranule A large granule on the Sun's surface including many smaller granules.

supernova A "new" star appearing in Earth's sky and lasting for a year or so before fading. Caused by the violent explosion of a star.

supernova (type I) The violent explosion of a star in which the spectrum contains no hydrogen lines.

supernova (type Ia) The explosion of a star caused by the collapse of a white dwarf that has gained mass from its binary companion and exceeds the Chandrasekhar limit.

supernova (type Ib) The explosion of a massive star that develops an iron core and collapses after it has lost its outer layers of hydrogen.

supernova (type II) The explosion of a massive star that develops an iron core and collapses.

supernova remnant The expanding shell of gas marking the site of a supernova explosion.

superwind Extremely rapid outflow of matter from giant and supergiant stars, analogous to the solar wind but with much greater mass loss rate.

synchrotron radiation Radiation emitted when high-speed electrons move through a magnetic field.

synodic period The period of rotation or revolution of a celestial body with respect to the Sun.

T association A large, loosely bound group of T Tauri stars.

T dwarf A very low-mass star at the bottom end of the main sequence with a cool surface and a low luminosity.

T Tauri star Young star surrounded by gas and dust contracting toward the main sequence.

temperature A measure of the velocity of random motions among the atoms or molecules in a material.

terminator The dividing line between daylight and darkness on a planet or moon.

Terrestrial planet Earth-like planet—small, dense, rocky.

theory A system of assumptions and principles applicable to a wide range of phenomena that have been repeatedly verified.

thermal energy The energy stored in an object as agitation among its atoms and molecules.

thermophile A type of microorganism that thrives in high-temperature environments.

tidal coupling The locking of the rotation of a body to its revolution around another body.

tidal heating The heating of a planet or satellite because of friction caused by tides.

tidal tail A long strand of gas, dust, and stars drawn out of a galaxy interacting gravitationally with another galaxy.

time dilation The slowing of moving clocks or clocks in strong gravitational fields.

top-down hypothesis The hypothesis that the Milky Way Galaxy formed by gravitational collapse of a single large spinning cloud of gas. This hypothesis is now considered inadequate to explain many observed characteristics of the galaxy.

total lunar eclipse A lunar eclipse in which the Moon completely enters Earth's dark shadow.

total solar eclipse A solar eclipse in which the Moon completely covers the bright surface of the Sun.

totality The period during a solar eclipse when the Sun's photosphere is completely hidden by the Moon, or the period during a lunar eclipse when the Moon is completely inside the umbra of Earth's shadow.

transit The passage of an extrasolar planet across the disk of its parent star as observed from Earth, partially blocking the light from the star and allowing detection and study of the planet.

transition The movement of an electron from one atomic orbit to another.

triple alpha process The nuclear fusion process that combines three helium nuclei (alpha particles) to make one carbon nucleus.

Trojan asteroid Small, rocky body caught in Jupiter's orbit at the Lagrange points, 60° ahead of and behind the planet.

turnoff point The point in an H–R diagram where a cluster's stars turn off the main sequence and move toward the red giant region, revealing the approximate age of the cluster.

21-cm radiation Radio emission produced by cold, low-density hydrogen in interstellar space.

ultraluminous infrared galaxy (ULIRG) A highly luminous galaxy so filled with dust that most of its energy escapes as infrared photons emitted by warmed dust.

ultraviolet (UV) radiation Electromagnetic radiation with wavelengths shorter than visible light but longer than X-rays.

umbra The region of a shadow that is totally shaded, especially the shadow of Earth or the Moon during an eclipse.

uncompressed density The density a planet would have if its gravity did not compress it.

unified model The attempt to explain the different kinds of active galaxies and quasars by a single model.

uniform circular motion The classical belief that the perfect heavens could move only by the combination of constant motion along circular orbits.

valley networks Dry drainage channels resembling streambeds found on Mars.

Van Allen belt One of the radiation belts of high-energy particles trapped in Earth's magnetosphere.

variable star A star whose brightness changes periodically.

velocity A rate of travel that specifies both speed and direction.

velocity dispersion method A method of finding a galaxy's mass by observing the range of velocities within the galaxy.

vernal equinox The place on the celestial sphere where the Sun crosses the celestial equator moving northward; also, the time of year when the Sun crosses this point, about March 21, and spring begins in the Northern Hemisphere.

vesicular A porous basalt rock formed by solidified lava with trapped bubbles.

violent motion In Aristotelian physics, motion other than natural motion that has to be sustained by a force. (See *natural motion*.)

visual binary system A binary star system in which the two stars are separately visible in the telescope.

void A region containing relatively few galaxies, part of the large-scale structure of the Universe.

water hole The interval of the radio spectrum between the 21-cm hydrogen radiation and the 18-cm OH radiation, likely wavelengths to use in the search for extraterrestrial life.

wavelength The distance between successive peaks or troughs of a wave; usually represented by l.

wavelength of maximum intensity (l_{max}) The wavelength at which a perfect radiator emits the maximum amount of energy; depends only on the object's temperature.

weak nuclear force One of the four forces of nature; the weak nuclear force is responsible for some forms of radioactive decay.

west point The point on the western horizon exactly halfway between the north point and the south point; exactly west.

white dwarf The remains of a dying star that has collapsed to the size of Earth and is slowly cooling off; at the lower left of the H–R diagram.

Widmanstätten pattern Bands in iron meteorites large crystals of nickel–iron alloys.

Wien's law The mathematical relation between the temperature of an ideal radiator (a blackbody) and the wavelength at which it radiates most intensely.

WIMP Weakly interacting massive particle, a hypothetical type of subatomic particle of which dark matter could be composed.

winter solstice The point on the celestial sphere where the Sun is farthest south; also, the time of year when the Sun passes this point, about December 22, and winter begins in the Northern Hemisphere.

X-ray Electromagnetic radiation with short wavelengths, high frequencies, and high photon energies, between gamma-rays and ultraviolet radiation on the electromagnetic spectrum.

X-ray burster An object that produces occasional X-ray flares. Thought to be caused by mass transfer in a closed binary star system.

Y dwarf A spectral type of brown dwarf with a temperature below 500 K.

Young Stellar Object (YSO) A forming star in a late protostellar evolutionary stage that has lost its obscuring cocoon of gas but is still contracting toward the main sequence.

Zeeman effect The splitting of spectral lines into multiple components when the atoms are in a magnetic field.

zenith The point directly overhead on the sky.

zero-age main sequence (ZAMS) The locus in the H–R diagram where stars first reach stability as hydrogen-burning stars.

zodiac The band around the sky centered on the ecliptic within which the planets move.

INDEX

Boldface page numbers indicate definitions of key terms.

A

A stars, 177
A0620-00, 302
Abell 1689, 391
Absolute magnitude and distance, 174–175
Absolute visual magnitude (M_v), 174–175
Absolute zero, **132**
Absorption spectrum, 136
Acceleration, **81**
Accretion disk, **271**
Accretion disk jets, 304
Achernar, 177, A-6
Achernar A, 183
Achromatic lens, **106**
Acrux, A-6
Active galactic nuclei (AGN), **362,** 370
Active galaxies, **362,** 364
Active optics, 113
Active regions, 155
Adaptive optics, **121**
Adenine (A), 412
Adhara, 184
Adventures of Huckleberry Finn, The (Twain), 217
Afterword, A-1 to A-2
AGILE (Astro Rivelatore Gamma a Immagini Leggero), 292
AGN (active galactic nuclei), **362,** 370
Airborne telescope, 116
Al Jabbar, 12
Al Magisti, 61
Alanine, 414
Aldebaran, 15, A-6
  giant (III), 184
  K5 star, 177
  luminosity, 175
  name, 13
  red giant, 246
  winter sky, 14
Alfonsine Tables, 58, 62, 64, 65
Algol, 189, 190, 270
Algol paradox, 270
ALH 84001, 419
Allen, Paul, 423
Allen Telescope Array (ATA), 423
ALMA (Atacama Large Millimeter/submillimeter Array), 101, 122, 264
Almagest, 58, 61
Alpha Aquilae, 183, 184
Alpha Canis Majoris, 15, 184. *See also* Sirius
α Cen A, A-7
α Cen B, A-7
Alpha Centauri, 14, 15, 172, 178
α CMa A, A-7
α CMa B, A-7
Alpha Eridani A, 183
Alpha Leonis A, 183
Alpha Lyrae, 15, 184. *See also* Vega
Alpha Orionis, 14, 15. *See also* Betelgeuse
Alpha particle, 248
Alpha Tauri, 15, 184. *See also* Aldebaran
Alpha Ursae Minoris, 15, 184
Alpha Virginis, 15
Alpher, Ralph, 383
Alpheratz, 13
Alt-azimuth mount, 113
Altair, 183, 184, A-6
Amalthea, A-8
Amino acids, 412, 414
Analogy, 380
Ancient Greece, 55–58
Ancient model of the Universe, 60–63
Andromeda, 13
Andromeda Galaxy, 311, 343, 350, 354
Angstrom (Å), **103,** A-5
Angström, Anders Jonas, 103
Angular diameter, 19, 40
Angular distance, 19
Angular momentum, **88**
Annular eclipse, **42**
Anomalocaris, 416
Ant-on-orange analogy, 393
Antennae, 353, 355
Anticoded, **421**
Anticoded message, 422
Antimatter, **385**
Aphelion, 23
Apogee, **42,** 87
Apparent visual magnitudes (m_v), 15, 16
Arc, 326
Arc minutes, 19
Arc seconds, 19
Archaea, **418**
Archaeoastronomy, **53**–55
Arcturus, 138, 139, A-6
Arecibo message, 421, 422
Arecibo Observatory, 289
Arecibo telescope, 115
Ariel, A-8
Ariel I, 117
Aristarchus, 57
Aristotle, 56–57, 60, 71, 79
Array detector, **118**
Association, **230,** 252, 316
Asterisms, **13**
Astro Rivelatore Gamma a Immagini Leggero (AGILE), 292
Astrobiology, 408–427
  biological evolution, 410
  Cambrian explosion, 416, 417
  chemical evolution, 414–415
  common misconceptions, 410, 416
  concept art, 412–413
  Copernican Principle, 424
  defined, **409**
  DNA, 410, 412–413
  Drake equation, 423–425
  extremophiles, 418
  geological time and evolution of life, 415–417
  habitable zone, 420
  intelligent life in the Universe, 420–425
  life in our Solar System, 418–420
  life on other planetary systems, 420
  Miller–Urey experiment, 411, 414
  mutations, 410
  natural selection, 410
  nature of life, 409–410
  number of technological civilizations per galaxy (table), 425
  origin of life on Earth, 411, 414–415
  physical bases of life, 409–410
  radio communication, 421–423
  search for extra-terrestrial intelligence (SETI), 422, 423
  UFOs, 420, 421
Astrology, 26

Astronomer's triangulation method, 171–172
Astronomia Nova (Kepler), 67
Astronomical cameras, 118–119
Astronomical constants, A-5
Astronomical instruments and techniques, 118–123
  adaptive optics, 121
  cameras and photometers, 118–119
  interferometry, 121–123
  spectrograph, 119–121
Astronomical interferometer, 122
Astronomical unit (AU), **3,** A-5
Astronomy
  defined, A-2
  gravity wave, 123
  high-energy, 117–118
  instruments and techniques, 118–123
  non-electromagnetic, 123
  origins. *See* Historical overview
  particle, 123
  why studied?, 7
ATA (Allen Telescope Array), 423
Atacama Large Millimeter/submillimeter Array (ALMA), 101, 122, 264
Atmospheric seeing, 108–109
Atmospheric window, **104**
Atom
  electron orbits, 129
  electrons/protons/neutrons, 128
  excitation, 131
  kinds, 129
Atomic spectra, 136–137. *See also* Spectrum (spectra)
Aurora borealis, 130
Auroras, 157, 159
Autumnal solstice, 22

B

B stars, 177
Baade, Walter, 287
Babcock model, **153**
Balmer-alpha line (Arcturus), 138
Balmer lines, 137, 175, 176
Balmer series, 135, 137
Balmer thermometer, 175
Barberini, Cardinal, 72
Barnard 86, 201, 203
Barnard's Star, A-7
Baronius, Cardinal, 73
Barred spiral galaxies, **340,** 356
Baryonic matter, 394, 396, 402
Baryons, 394
Basic scientific data, 194
Bayer, Johann, 14
Bayer designations, 14, 15
Becklin, Eric, 323
Becklin-Neugebauer (BN), 225
Bell, Jocelyn, 288
Bellarmine, Cardinal, 72
Berkeley SETI team, 423
Beta Orionis, 14, 15, 184. *See also* Rigel
Beta Persei, 189, 270. *See also* Algol
Betelgeuse, 15, A-6
  color, 132, 134
  cool, red supergiant, 246
  gas and dust, 212
  luminosity class, 184
  M2 star, 177, 178
  name, 13
  winter sky, 14
Bethe, Hans, 221
BICEP2 far-infrared telescope, 403
Big Bang, **381**
Big Bang Theory, 381–388, 400–401
Big Dipper, 12, 13, 173, 338
Big rip, **400**
Binary pulsars, 293–295
Binary stars, **185**–189, 270–272
Binding energy, **129**
Biological evolution, **410**
Bipolar flows, 229
Birth line, **226**
"Black cloud," 201
Black dwarfs, **269**
Black Eye Galaxy, 355
Black hole, 297–302
  binaries, 302
  common misconceptions, 299
  defined, **298**
  escape velocity, 297–298
  gravitational redshift, 299
  Schwarzschild, 298–299
  searching for, 299–302
  supermassive, 324
  time dilation, 299
Black hole binaries, 302
Black hole x-ray nova, 302
Black Widow pulsar, 296
Blackbody radiation, **133**–134
Blue dwarf galaxies, **356**
Blueshift, 138, **139,** 353
BN (Becklin-Neugebauer), 225
Bok, Bart, 224
Bok globules, 224
Boltzmann, Ludwig, 133
Bottom-up hypothesis, **331**
Bow stock, 211, 212
Brahe, Tycho, 64–66
Brecht, Bertolt, 73
Bright-line spectra, 136
Brightest stars, 193, A-6
Brown dwarfs, 178, **243**
Bruno, Giordano, 73
Buckyballs, 204

C

C153, 356
Calibration, **313,** 314
Callisto, A-8
Cambrian explosion, **416,** 417
Cane, 326
Canis Major, 192
Canis Major Dwarf Galaxy (CMa Dwarf), 350, 351, 354
Canopus, 177, A-6
Capella, A-6
Carbon, nitrogen, and oxygen (CNO) abundances, 150
Carbon deflagration, **277**
Carbon fusion, 251
Cartwheel Galaxy, 353
Cassegrain focus, 112
Cassiopeia A (Cas A), 279
Cat's Eye Nebula, 266, 267
Cause and effect, 83
CCD (charge-coupled device), **118**
Celestial bear, 12
Celestial equator, 18
Celestial sphere, 17–20
Cell reproduction by division, 413
Celsius, Anders, A-4
Celsius temperature scale, 132, A-4
Centaurus A, 366
Center for High Angular Resolution Astronomy (CHARA) array, 123, 183
Center of mass, 87
Centigrade temperature scale, A-4
Central bulge, 316, **317,** 346
Cepheid variable stars, 253, 256–258, 337, 342, 343
Cepheus, 256
CGRO (Compton Gamma Ray Observatory), 118, 304
Chaco Canyon, 54
Chains of inference, 186
Chandra X-ray Observatory, 117–118
  black hole candidates (M31), 302
  Cas A, 279
  deep-field images, 337
  hot gas and dark matter, 400
  M101, 117–118, 120
  Milky Way Galaxy, 324
  Sgr A*, 327
Chandrasekhar, Subrahmanyan, 118, 269
Chandrasekhar limit, **269**

CHARA (Center for High Angular Resolution Astronomy) array, 123, 183
Charge-coupled device (CCD), **118**
Charon, A-8
Chemical evolution, **414**–415
Chickadees, 301
Choldera, 83
Chort, 15
Chromatic aberration, **106**
Chromosomes, 413
Chromosphere, **43, 144,** 146–147
Circular velocity, 84, 86
Circumpolar constellations, 19
Classification, 339
Clone, 413
Closed orbits, 87
Closed universe, 390, 393
Cluster method, **346**
Clusters of galaxies, 349–350
CMa Dwarf (Canis Major Dwarf Galaxy), 350, 351, 354
CME (coronal mass ejection), 159
CNO (carbon, nitrogen, and oxygen) abundances, 150
CNO cycle, **232**–233
Coalsack Nebula, 199
COBE satellite, 384, 402
Cocoon nebula, **220**
Cold dark matter, **396**
Columbus, Christopher, 57
Coma cluster, 347, 348
Common misconceptions
  atoms, 129
  Big Bang, 382
  black hole, 299
  Columbus, Christopher, 57
  cosmological redshifts, 389
  evolution of life from simple beginnings, 416
  explosions in space, 402
  Galileo, 70
  gravity, 85
  light-year, 4
  Moon, 36, 41
  natural selection, 410
  nuclear fusion in the Sun, **163**
  Polaris, 20
  radiation, 102
  radio waves, 103
  seasons, 24
  solar eclipse, 44
  standing raw egg on end, 24
  stars, 5, 20, 253
  Sun's core, 233
  telescope, 110
  terminology (Galaxy/Solar System/Universe), 6
  thermal energy, heat, temperature, 132
  tides, 89
  Universe, 378–379, 380
  vacuum of space, 200
  violent events and lack of noise, 274
Compact object
  accretion disk jets, 304
  defined, **268**
  gamma-ray bursts (GRBs), 304–306
  jets of energy, 303–304
Comparison spectrum, **119**
Compton Gamma Ray Observatory (CGRO), 118, 304
Concept art
  ancient model of universe, 60–63
  atomic spectra, 136–137
  cosmic jets and radio lobes, 366–367
  cycle of the seasons, 22–23
  DNA, 412–413
  family of stars, 192–193
  galaxy classification, 340–341
  interacting galaxies, 352–353
  lighthouse model of pulsars, 290–291
  Milky Way Galaxy, 326–327
  nebulae, 202–203
  optical telescopes, 112–113
  orbits, 86–87
  phases of Moon, 34–35
  planetary nebulae, 266–267
  prostellar disks, 228–229
  sky around you, 18–19
  solar activity and Sun–Earth connection, 158–159
  space-time, 392–393
  star clusters and stellar evolution, 254–255
  star formation in Orion Nebula, 224–225
  sunspots and solar magnetic cycle, 154–155
  white dwarfs, 266–267
  young stellar objects, 228–229
Conduction, 231
Confirmation and consolidation, 157
Conservation of angular momentum, 288
Consolidation, 157
Constellations, **12**–13
Contagious star formation, 222–223
Continuous spectrum, 136
Convection, **145,** 231–232
Convective cores, 233, 234
Convective zone, **164**
Cool clouds, 208
Cooling lines, **220**
Copernican Principle, **424**
Copernicus, Nicolaus, 58–63, 424
Corona, **43, 144,** 147–149
Coronagraph, **147**
Coronal gas, 209, **210**–211
Coronal hole, 159
Coronal mass ejection (CME), 159
Cosmic jets, 366–367
Cosmic microwave background radiation, **383**–384
Cosmic rays, **123**
Cosmological constant, **397**
Cosmological principle, **389**
Cosmological redshifts, 365, 389, 392
Cosmologists, **379**
Cosmology, 377–407
  acceleration and dark energy, 397–399
  age of Universe, 382
  ant-on-orange analogy, 393
  baryonic/nonbaryonic matter, 394, 396, 402–403
  Big Bang Theory, 381–388, 400–401
  big rip, 400
  closed, flat, or open universe, 393, 402
  common misconceptions, 378–379, 380, 382, 389, 402
  concept art, 392–393
  conclusions (fitting it all together), 401–403
  cosmic background radiation, 383–384
  cosmic element building, 385, 386
  cosmic expansion, 380–381
  cosmological constant, 397
  cosmological principle, 389
  dark age, 387
  dark energy, 398, 403
  dark energy and evolution of Universe, 399–400
  dark matter, 391, 394, 396, 403
  defined, **378**
  edge-center problem, 378–379
  flatness problem, 400
  grand unified theories (GUTs), 401
  horizon problem, 400
  Hubble time, 382, 391
  inflationary hypothesis, 400–401
  isotropic, homogeneous universe, 388
  large-scale structure, 394–396
  nucleosynthesis, 385–386
  observable universe, 380
  Olbers's paradox, 379
  ordinary matter and dark matter, 391, 394
  overview (cosmology pyramid), 378
  photon and particle soup, 385
  quintessence, 398
  raisin bread analogy, 381
  recombination and re-ionization, 386–388
  redshifts, 389, 392
  shape of space, 389–390, 392–393, 402

Cosmology pyramid, 378
Coulomb, Charles-Augustin de, 129
Coulomb barrier, **164,** 232
Coulomb force, **129,** 164, 232
CP 1919, 288
Crab Nebula, 278, 289, 291
Crab Nebula pulsar, 289, 292
Critical density, **390**
Curiosity rover, 419
Curtis, Heber, 337
Curvature of space-time, 94–97
Cygnus, 208
Cygnus A, 366
Cygnus Loop, 279
Cygnus X-1, 301, 302
Cytosine (C), 412

D

Dark age, **387**
Dark energy, **398,** 403
Dark-line spectra, 136
Dark matter, **319,** 347–348, 391, 394, 396, 403
Dark nebulae, 203
Darwin, Charles, 339
Darwin's theory of evolution, 69
Davis, Raymond, Jr., 165
Davis neutrino experiment, 165, 166
De Revolutionibus (Copernicus), 59, 62–63, 72
De Stella Nova (Bahe), 65
Death of stars, 262–285
  binary stars, 270–272
  concept art, 266–267
  end of Earth, 282
  low-mass stars. *See* Low-mass stars
  nova explosions, 271–272
  nuclear fusion in massive stars, 273–274
  supernova explosions. *See* Supernova explosions
Deep fields, 337
Deferent, 61
Degenerate matter, 247–**248**
Deimos, A-8
Delta Cephei, 253, 256
Dense core, **218**
Deoxyribonucleic acid (DNA), 410, **412**–413
Deuterium, **163**
Dialogo Sopra i Due Massimi Sistemi del Mondo (Galileo), 72, 73
Diamond ring effect, **43**
Dicke, Robert, 383
Differential rotation, **153,** 318
Diffraction fringes, **108**
Digges, Thomas, 379
Dione, A-8
Discourse and Mathematical Demonstrations Concerning Two New Sciences, Relating to Mechanics and to Local Motion (Galileo), 81
Disk component, **315**–317
Disk stars, 316, 317
Distance, 392
Distance indicators, **342**
Distance ladder, 343
Distance scale, **343**
DNA (deoxyribonucleic acid), 410, **412**–413
Doppler effect, **138**–139, 345
Doppler shift, 139
Doppler velocities, 139–140
Double-exhaust model, **366**
Double-lobed radio galaxies, **363**
Drake, Frank, 423
Drake equation, **423**–425
Dwarf, 179, 181, 184
Dwarf ellipticals, 339, 356
Dwarf irregulars, 339
Dwarf planet, **4**
Dynamo effect, **152**

E

E-ELT (European Extremely Large Telescope), 114
E0 to E7 galaxies, 340
Earth
  astronomical influences on Earth's climate, 25–26
  end of, 282
  moon. *See* Moon (Earth)
  orbit, 24
  orbital properties, A-7
  physical properties, A-7
  solar activity and Sun–Earth connection, 158–159
  supernova explosions, 281
East point, 18
Eccentricity *(e),* **67**
Eccentrics, **57**
Eclipse
  lunar, 36–38
  predictions, 45–48
  solar, 38–45
Eclipse prediction, 45–48
Eclipse season, **46**
Eclipse year, **46**
Eclipsing binary system, **187,** 189
Ecliptic, **21**
Eddington, Arthur, 221, 239
Edge-center problem, 378–379
Egg Nebula, 267
Einstein, Albert, 78, 91, 92–97, 298, 348, 390, 397
Einstein@Home project, 289
Einstein's energy-mass equivalence, 94
Electromagnetic radiation, **102**
Electron, **128**
Electroweak force, 401
Elephant Trunk, 228
Eliot, T. S., A-2
eLISA (Evolved Laser Interferometry Space Antenna), 123
Ellipse, **67**
Elliptical galaxies, 339, **340,** 344, 346, 354
Emission nebula, 202, 203
Emission spectrum, 136
Empirical, **68**
Enceladus, A-8
End of Earth, 282
Energy, **88**
Energy conservation, 239, 240
Energy levels, **131**
Energy transport, **231**–232, 240
Enzymes, 412, 413
Epicycle, 61
Epsilon Canis Majoris, 184
ϵ Eri, A-7
Equant, 61
Equatorial mount, 113
Equinoxes, 22, 24
Equivalence principle, 94
Eratosthenes, 57
Escape velocity, 85, 87
Eskimo Nebula, 128
ESO 510-G13, 336
ESO Very Large Telescope (VLT), 111, 114, 122
Eta Carinae, 238, 242
Eudoxus of Cnidus, 56
Europa, 418, A-8
European Extremely Large Telescope (E-ELT), 114
European Very Large Telescope (VLT), 111, 114, 122
Evening star, **25**
Event horizon, **298,** 299
Event Horizon Telescope, 324
Evolved Laser Interferometry Space Antenna (eLISA), 123
Excited atom, **131**
Expanding universe, **380**
Experimentalist, 70
Extrasolar planet, 5

Extraterrestrial radio signals, 422
Extreme Population I stars, 328
Extreme Population II stars, 328
Extremophile, **418**
Eyepiece, **106**

F

F stars, 177
Fahrenheit, Gabriel Daniel, A-4
Fahrenheit temperature scale, 132, A-4
False-color images, **119**
Famous stars, 14, 15
Farthest galaxies, 356–357
FAST, 115
Fermi Gamma-Ray Space Telescope (Fermi), 118, 292, 305
Field, **84**
Field of view, **2**–6
Filaments, 158, **395**
Filtergram, **146**
First-magnitude stars, 14
First postulate of relativity, 92, 93
First principles, **56**
Flares, 159
Flat universe, 390, 393
Flatness problem, **400**
Flocculent galaxy, **322**
Floppy mirror, 113
Flux, **15, 174**
Flux ratio, 15, 16
Focal length, **106**
Forbidden lines, **206**
Formation and structure of stars, 216–237
  birth line, 226
  CNO cycle, 232–233
  concept art, 224–225, 228–229
  contagious star formation, 222–223
  energy transport, 231–232
  free-fall collapse, 219
  heating by contraction, 219–220
  hydrostatic equilibrium, 230, 231
  interstellar medium (ISM), 212, 217–221
  observing star formation, 222, 224–225
  Orion Nebula, 221–225
  pressure–temperature thermostat, 233–234
  proton–proton chain, 232
  protostars, 220
  protostellar disks, 226, 228–229
  star birth in giant molecular clouds, 217–219
  star-formation pillars, 222
  stellar energy, 232–234
  stellar structure, 230–232
  young stellar associations, 230
  young stellar objects (YSOs), 226, 228–229
Fornax A, 364, 365
47 Tucanae, 254
Four forces of nature, 161–162
Four laws of stellar structure, 239, 240
4U 1820-30, 295, 296
Fraud in science, 301
Free-fall collapse, **219**
Frequency, **102**
Frost, Robert, 102
FU Orionis phenomenon (FU Ori), **227**
Fuller, Buckminster, 204
Fusion reactions, 162–164

G

G stars, 177
G112-0.3, 293
Gaia spacecraft, 173
Galactic cannibalism, **352,** 353
Galactic center, 315, 316
Galactic corona, 318–**319**
Galactic fountain, **329**
Galactic plane, 316, 317
Galaxies, **5,** 335–360
  active, 362
  active galactic nuclei (AGN), 362, 370
  central bulge, 346
  classification, 338, 340–341
  clusters, 349–350
  concept art, 340–341, 352–353, 366–367
  dark matter, 347–348
  diameter and luminosity, 344, 346
  distance, 342–343
  double-lobed radio sources, 363–365
  farthest, 356–357
  gravitational lensing, 348, 349
  Hubble law, 343–344
  interactions (collisions/mergers), 350–354
  mass, 345–346
  origin/evolution, 354–356
  Seyfert, 362–363
  supermassive black holes, 346–347. *See also* Supermassive black holes
Galaxy classification, 338, 340–341
Galaxy clusters, 349–350
Galaxy collisions, 350–354
Galaxy mergers, 350–354
Galaxy Zoo, 115
Galilean moons, **71**
Galilei, Galileo, 53, 70–73, 79–81, 105, 151, 311
Galileo's telescopes, 70
Gamma-ray burst (GRB), **304**–306
Gamma-ray space telescope, 118
Gamma rays, **104, 163**
Gamow, George, 383
Ganymede, A-8
Gas-stars-gas cycle, 212, 213
Gathering light, 107
Gemini-North optical telescope, 107
Gene, 413
General theory of relativity, 94, 294
Geocentric universe, **55,** 60
Geological time and evolution of life, 415–417
Geosynchronous satellite, 86
Germ theory of disease, 69
Giant, **181,** 182, 195, 246–247
Giant ellipticals, 344
Giant Magellan Telescope (GMT), 114
Giant molecular cloud, **210,** 217–219
Global Oscillation Network Group (GONG), 150
Globular cluster, 253, 254, 313, 315, 330, 342
GMT (Giant Magellan Telescope), 114
GONG (Global Oscillation Network Group), 150
Goodricke, John, 253
GOODS deep fields, 399
GOODS (Great Observatories Origins Deep Survey), 337
Google Sky, 115
Gran Telescopio Canarias (GTC), 111, 114
Grand-design galaxy, **322**
Grand Prismatic Spring, 418
Grand unified theories (GUTs), **401**
Granulation, **145,** 146
Granules, 145, 146
Grating, **119**
Gravitational lensing, **348,** 349
Gravitational radiation, 96–97, **294**
Gravitational redshift, **299**
Gravity, 78–100
  common misconception, 85
  curvature of space-time, 94–97
  Einstein, Albert, 92–97
  Einstein's energy-mass equivalence, 94
  equivalence principle, 94
  Galileo's observations of motion, 79–81
  general theory of relativity, 94
  gravitational radiation, 96–97
  mutual gravitation, 82–84
  Newton's law of gravitation, 84
  Newton's laws of motion, 81–82

Gravity (*continued*)
Newton's version of Kepler's third law, 88–89
orbital motion, 84–89
postulates of relativity, 93
Principia (Newton), 91
special relativity, 92–94
tides and tidal forces, 89–91
Gravity according to general relativity, 94
Gravity wave astronomy, 123
GRB (gamma-ray burst), **304**–306
Great Nebula in Orion, 221–225
Great Observatories Origins Deep Survey (GOODS), 337
Great Rift, 203
Great Square of Pegasus, 13
Greek alphabet, A-9
Greek-letter designation, 14
Gregory XV, Pope, 72
Grindstone universe, 312
GRO J1655-40, 302, 303
Ground-based telescope, 114
Ground state, **131**
Guanine (G), 412
Guth, Alan, 401
GUTs (grand unified theories), **401**

H

H-alpha, 158
H-alpha filtergram, 147
H-beta, 158
H-gamma, 158
H I clouds, 208, 209
H II region, 202, 203
H–R (Hertzsprung–Russell) diagram, **180**–181, 182, 191
Habitable zone, **420**
Hadar, A-6
Halo, **317**
Halo population star, 328
Halo stars, 317
"Hand of God," 225
Hanny's Voorwerp, 371, 372
Harmonices Mundi (Kepler), 67
Haro, Guillermo, 229
Hawking, Stephen, 298
HD 12545, 161
HD 80715, 188
Heat, **132**
Heavy-element fusion, 274
Heinlein, Robert, 263
Heliocentric universe, **55,** 58–59, 75
Heliopause, **149**
Helioseismology, **150,** 151
Helium flash, **248,** 249
Helium fusion, 248–251
Helix Nebula, 266
Henize 206, 219
Herbig, George, 229
Herbig-Haro objects, 229
Hercules galaxy cluster, 350
Hercules X-1, **294,** 295, 301
Herman, Robert, 383
Herschel, Caroline, 312
Herschel, William, 312, 337
Herschel 3-meter infrared space telescope, 117
Hertzsprung, Ejnar, 180
Hertzsprung–Russell (H–R) diagram, **180**–181, 182, 191
HH30, 227
Hickson Compact Group 87, 344
Higgs boson, 92
High-energy astronomy, 117–118
High-mass stars
hydrogen fusion, 233
mass loss, 242
neutrino-antineutrino particle pairs, 274
nuclear fusion, 273–274
stability, 242, 243
High-precision spectroscopic analysis, 160
Himalia, A-8
Hinode spacecraft, 117, 148
Hipparchus, 14, 57
Hipparcos satellite, 173, 330
Historical overview, 52–77
ancient Greece, 55–58
ancient model of Universe, 60–63
archaeoastronomy, 53–55
Aristotle, 56–57
Brahe, Tycho, 64–66
Copernicus, Nicolaus, 58–63
De Revolutionibus (Copernicus), 59, 62–63
Galileo, 70–73
geocentric universe, 55
heliocentric universe, 58–59, 75
Kepler, Johannes, 66–69
Kepler's three laws of planetary motion, 67–68
perfection of the heavens, 56
Ptolemaic universe, 58
Rudolphine Tables (Kepler), 68–69
timeline (99 years that revolutionized astronomy), 74
Historical supernovae, 278–280
Homogeneous, **389**
Horizon, 18
Horizon problem, **400**
Horizontal branch, 255
Horoscope, **25**
Horsehead Nebula, 204
Hot dark matter, **396**
Hot spots, **366**
Hour Glass Nebula, 267
How Do We Know? boxes
basic scientific data, 194
calibration, 314
cause and effect, 83
chains of inference, 186
classification in science, 339
confirmation and consolidation, 157
Copernican Principle, 424
fraud in science, 301
hypotheses, theories, and laws, 69
mathematical models, 241
nature as processes, 323
pseudoscience, 26
quantum mechanics, 130
reasoning by analogy, 380
resolution and precision, 109
science: a system of knowledge, 388
scientific arguments, 28
scientific confidence, 165
scientific imagination, 39
scientific method, 8
scientific models, 17
scientific revolutions, 64
searching for natural causes, 268
separating facts from hypotheses, 209
statistical evidence, 364
testing hypothesis by prediction, 92
theories and proof, 221
wishing doesn't make it so, 403
HST, 116. *See also Hubble Space Telescope*
Hubble, Edwin, 116, 337, 339, 344, 354, 380, 392, 397
Hubble constant, **344**
Hubble Deep Field, 337, 338
Hubble Deep Field-South, 424
Hubble law, **344,** 380, 392
Hubble Space Telescope
Antennae, 353
black holes, 369
Cepheids, 342
deep-field images, 337
distances to galaxies, 397
galaxy, core of, 353
gamma-ray burst (GRB), 304
HH30, 227
Hubble time, 391
M101, 120
Mars, 109
Milky Way Galaxy, 324, 327
Nova T Pyxidis, 272
Orion Nebula, 225
overview, 116–117

planetary nebulae, 267
RX J185635-3754, 291
Sgr A*, 327
space shuttle, 117
type 1a supernova, 343
Hubble time, **382,** 391
Hulse, Russell, 293, 294
Humason, Milton, 344
Hveen, 65
Hyades Star Cluster, 254
Hydra, A-8
Hydra cluster, 380
Hydrogen Balmer lines, 176, 177
Hydrogen fusion, 162–164, 246, 248
Hydrostatic equilibrium, **230,** 231, 240, 242
Hyperion, A-8
Hypernova, **305**
Hypothesis
defined, **69**
facts, compared, 209
prediction, 92
scientific method, 8
theory, and, 221

I

Iapetus, A-8
IAU (International Astronomical Union), **13**
IBEX (Interstellar Boundary Explorer) spacecraft, 149, 211
IC 2163, 341
IC 2497, 371, 372
IC 4182, 341
Ice, 268
In-depth discussion. *See* Concept art
Inertia, **81**
Inflation-modified Big Bang Theory, 400–401
Inflationary universe, **400**–401
Infrared (IR) radiation, **103**–104
Inner Lagrange point, **270**
Instability strip, **256**
INTEGRAL (INTErnational Gamma-Ray Astrophysics Laboratory), 118
Intercloud medium, **208**–209
Interface Region Imaging Spectrograph (IRIS) space telescope, 148
Interferometry, **121**–123
Intergalactic medium, **365**
Intermediate Population I stars, 328
Internal Constitution of the Stars, The (Eddington), 239
International Astronomical Union (IAU), **13**
INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL), 118
International Space Station, 348
Interstellar absorption lines, **205**–206
Interstellar Boundary Explorer (IBEX) spacecraft, 149, 211
Interstellar dust, **201,** 204
Interstellar emission lines, **206**
Interstellar extinction, **201**
Interstellar medium (ISM), 199–215
absorption lines, 205–206
composition, 208–211
concept art, 202–203
cool clouds, 208
coronal gas, 209, 210–211
defined, **200**
emission lines, 206
extinction and reddening, 201, 204
gas-stars-gas cycle, 212, 213
H I clouds, 208, 209
infrared radiation from ISM dust, 204–205
intercloud medium, 208–209
interstellar dust, 201, 204
molecular clouds, 209–210
nebulae, 200–201, 202–203
Solar System material, 211
star formation, 212, 217–221
21-cm radiation, 206–207
Interstellar reddening, **201,** 204
Interval, 392
Intrinsic brightness, **174**
Intrinsic variables, **253**
Inverse square law, **82,** 83, 174
Invisible light, 103
Io, A-8
Ion, **129**
Ionization, **129**
IR (infrared) radiation, **103**–104
IRIS (Interface Region Imaging Spectrograph) space telescope, 148
Iron-60, 281
Irr galaxies, 341
Irregular galaxies, **341,** 344, 346, 356
Island universe, **337**
ISM. *See* Interstellar medium (ISM)
Isotope, **129**
Isotropic, **389**

J

James Webb Space Telescope (JWST), 117
Janus, A-8
Jewel Box cluster, 254
John Paul II, Pope, 73
Joule (J), **88,** A-3
Jupiter
moons, A-8
orbital properties, A-7
physical properties, A-7
JWST (James Webb Space Telescope), 117

K

K stars, 177, 178
Kant, Immanuel, 337
Keck telescopes, 113, 114, 123
Kelvin temperature scale, **132,** A-4
Kepler, Johannes, 66–69
Kepler space observatory, 117
Kepler space telescope, 160
Keplerian motion, 318
Kepler's first law, 67, 68, 85
Kepler's second law, 67, 68, 87, 88
Kepler's supernova of 1604, 278, 280
Kepler's third law, 67, 68, 88, 369
Kepler's third law for stellar masses, 185
Kepler's three laws of planetary motion, 67–68
Kerr, Roy, 298
Kiloparsec (kpc), **312,** A-5
Kinetic energy, 88
Kirchhoff, Gustav, 136
Kirchhoff's laws, 136
Kirchhoff's second law, 136, 146
Kirchhoff's third law, 136, 145
Kitt Peak National Observatory, 112
Kleinmann-Low (KL), 225
Koch, Robert, 83

L

L dwarfs, 178, 179
Lacaille 9352, A-7
Lagoon Nebula, 203
Lagrange points, **270**
Lalande 21185, A-7
λ Orionis, 177
Laozi, 2
Large Binocular Telescope (LBT), 111, 114, 123
Large Magellanic Cloud, 280, 341
Large-scale structure, **394**–396
Large Synoptic Survey Telescope (LSST), 114, 115
Laser guide star, **121**
Laser Interferometer Gravitational Wave Observatory (LIGO), 96–97, 123, 294
Law of energy conservation, **239,** 240
Law of hydrostatic equilibrium, 230, 231
Law of mass conservation, **239,** 240

LBT (Large Binocular Telescope), 111, 114, 123
Leavitt, Henrietta, 313
Lenticular galaxies, **341**
Leo 1, 340
Life expectancy of stars, 245
Life on other worlds, 408. *See also* Astrobiology
Light
  electromagnetic spectrum, 103–104
  matter, and, 131–134
  visible spectrum, 104
  waves and particles, 102–103
Light curve, **188,** 253
Light-gathering power, **107**–108
Light pollution, **110**
Light-year (ly), **4,** A-5
Lighthouse model, **289,** 290–291
LIGO (Laser Interferometer Gravitational Wave Observatory), 96–97, 123, 294
Limb, **146**
Limb darkening, **146**
Linde, Andrei, 401
Linear diameter, 40
Lines of nodes, **46**
Little Ice Age, 155, 160
LMCX-3, 302
Lobachevsky, Nikolai, 393
Local bubble, **211**
Local Group, **349**–351
Local Supercluster, 395
Look-back time, **343,** 382
Low-mass stars
  black dwarfs, 243, 269
  Chandrasekhar limit, 269
  concept art, 266–267
  convective core, 236
  hydrogen fusion, 233
  mass loss and stellar evolution, 2642–65
  medium-mass stars, 264
  planetary nebulae, 265, 266–267
  red dwarfs, 263
  white dwarfs, 265–269
Lower-main-sequence red dwarfs, 191, 192
Lower-main-sequence stars, 263. *See also* Low-mass stars
LSST (Large Synoptic Survey Telescope), 114, 115
Luminosity
  defined, **175**
  galaxies, 344, 346
  main-sequence stars, A-6
  mass-luminosity relation, 191, 194
  period-luminosity relation, 253
  radius and temperature, 179
  rarity of luminous stars, 193
  spectral classification, 183–184
Luminosity classes, **183**
Luminosity-radius-temperature equation, 179
Luminosity spectral classification, 183–184
Luminous stars, 193
Lunar eclipse, **36**–38
Lunar phase, **33**
Luther, Martin, 59, 74
Luyten 726-8A, A-7
Luyten 726-8B, A-7
ly (light-year), **4,** A-5
Lyman lines, 137
Lyman series, 137

M

M stars, 177, 178
M1-67, 269
M2-9, 267
M19, 313
M31, 354
M52, 313
M64, 353
M67, 255
M82, 281
M83, 336
M87, 340, 349, 361
M100, 342
M101, 120
M104, 335
Magellanic Clouds, 341
Magnetars, **305**
Magnetic carpet, **148**
Magnifying power, **110**
Magnitude scale, **14**
Main sequence, **181,** 182
Main-sequence mass-luminosity relation, 242
Main-sequence stars, 239–245
  aging, 243–244
  density, 195
  life expectancy, 245
  lower end of main sequence. *See* Low-mass stars
  mass-luminosity relation, 242
  ordered by mass, 191
  properties, A-6
  reason for, 241–242
  upper end of main sequence. *See* High-mass stars
  years on main sequence, 244
  zero-age main sequence (ZAMS), 244
Mars
  moons, A-8
  orbital properties, A-7
  physical properties, A-7
Mass
  defined, **81**
  Einstein's energy-mass equivalence, 94
  galaxies, 345–346
  high-mass stars. *See* High-mass stars
  Kepler's third law for stellar masses, 185
  low-mass stars. *See* Low-mass stars
  main sequence, 191
  medium-mass stars, 264
  Milky Way Galaxy, 317–319
  neutron stars, 287
  nuclear reactions in massive stars, 252
  stars, 185–189, 191
  supermassive black holes, 346–347
Mass conservation, 239, 240
Mass-luminosity relation, **191,** 194, 242
Mass transfer, 270–271
Mathematical models, 241
Maunder, E. Walter, 154
Maunder butterfly diagram, 154
Maunder minimum, 155
Maxwell, James Clerk, 400
Medium-mass stars, 264
Megaparsec (Mpc), **342,** A-5
Mendel, Gregor, 8
Menzel 3, 267
Mercury, 25, A-7
Meridional flow, **156**
Messier 16, 229
Messier 31, 311, 337
Messier 42, 221. *See also* Orion Nebula
Messier 51, 276
Metals, **325**
Metastable levels, **206**
Meteor shower, A-9
Meteorological statistics, 364
Methematike Syntaxis (Ptolemy), 61
Metric prefixes, A-3
Metric system, A-3
Mice, 352
Microspheres, 415
Microwaves, **104**
Midsummer's Night's Dream, A (Shakespeare), 24
Milankovitch, Milutin, 26
Milankovitch cycles, 26–28
Milankovitch hypothesis, **26**
Milky Way, 5
Milky Way Galaxy, **5,** 6, 310–334
  age, 329–330
  bottom-up hypothesis, 331
  cannibal galaxy, 351
  central bulge, 316, 317
  components, 315–317
  concept art, 326–327
  differential rotation, 318
  discovery, 311–315

disk component, 315–317
element-building process, 325
formation, 330–332
galactic center, 315, 316
galactic corona, 318–319
galactic fountain, 329
galactic plane, 316, 317
halo, 317
mass, 317–319
nucleus, 323
origin/history, 325–332
Population I/II stars, 328
rotation curve, 318
SBbc galaxy, 340
self-sustaining star formation, 322–323
Sgr A*, 323, 324, 326–327
size, 312–315
spherical component, 317
spiral arms, 319–323
spiral density wave theory, 321
stellar populations, 328
structure, 315–319
supermassive black hole, 324
top-down hypothesis, 330–331
Miller, Stanley, 411, 414
Miller–Urey experiment, **411,** 414
Millisecond pulsars, **295**
Mimas, A-8
Miranda, A-8
Misconceptions. *See* Common misconceptions
Mitchell, John, 298
Molecular clouds, **209**–210
Molecule, **129**
Momentum, **81**
Monolithic collapse hypothesis, 330
Moon (Earth), A-8
common misconceptions, 36, 41
lunar eclipse, 36–38
orbital motion, 33
phases, 33–36
Moon phases, 33–36
Moons (principal satellites), A-8. *See also* Moon (Earth)
Morning star, **25**
Mu Cassiopeiae, 184
Multicellular, **416**
Murchison meteorite, 415
Mutation, **410**
Mutual gravitation, 82–84
Mysterium Cosmographicum (Kepler), 66

N

N11Br, 223
N44, 219
N90, 210
Nadir, 18
Nanometer (nm), **103**
Narratio Prima (Rheticus), 59
NASA SETI project, 423
National Radio Astronomy Observatory, 107
Natural law, **69**
Natural motions, **79**
Natural selection, **410**
Nature as processes, 323
Neap tides, **90**
Nearest stars, 193, A-7
Nebulae, 200–201, 202–203
Neon sign, 132
Neptune
moons, A-8
orbital properties, A-7
physical properties, A-7
Nereid, A-8
Neutrino-antineutrino particle pairs, 274
Neutrinos, **163,** 164–166
Neutron, **128**
Neutron star x-ray nova, 302
Neutron stars
defined, **287**
extrasolar planets (pulsar planets), 296–297
mass, 287
pulsars. *See* Pulsars
size, 288
spin, 288
x-ray bursters, 295
New Horizons spacecraft, 420
New moon, 34
Newgrange, 53
Newton (N), A-3
Newton, Isaac, 69, 78, 79, 82–84, 91, 106, 119, A-3
Newtonian focus, 112
Newton's first law of motion, 81, 82, 86
Newton's law of gravitation, 84
Newton's laws of motion, 81–82
Newton's second law of motion, 81–82, 83
Newton's third law of motion, 82, 83
Newton's version of Kepler's third law, 88–89, 185
NGC 346, 222
NGC 383, 367
NGC 602, 210, 222
NGC 604, 210
NGC 891, 119
NGC 1265, 367
NGC 1275, 364, 373
NGC 1316, 364, 365
NGC 1365, 340
NGC 1410, 362
NGC 1569, 355
NGC 1672, 362
NGC 1705, 355
NGC 1999, 224
NGC 2035, 217, 222
NGC 2207, 341
NGC 2264, 228, 255
NGC 2359, 242
NGC 2392, 128
NGC 2997, 340
NGC 2998, 345
NGC 3341a, 336
NGC 3341b, 336
NGC 3623, 340
NGC 4013, 341
NGC 4038, 353
NGC 4039, 353
NGC 4261, 369
NGC 4414, 336
NGC 5128, 352, 366
NGC 6520, 203
NGC 6624, 296
NGC 7251, 353
NGC 7469, 362
NGC 7674, 363
Nix, A-8
NM64, 355
Nodes, **45**
Non-electromagnetic astronomy, 123
Nonbaryonic matter, **394,** 403
North celestial pole, 18, 19, 21
North point, 18
Northern Hemisphere star charts, A-11, A-12 to A-13
Northern lights, 130, 157
Notebooks of Lazarus Long, The (Heinlein), 263
Nova, **271**
Nova Cygni 1975, 272
Nova explosions, 271–272
Nova T Pyxidis, 272
Nuclear fission, **162,** 163
Nuclear fusion, **162**
Nuclear fusion in massive stars, 273–274
Nuclear particle physicists, 268
Nucleosynthesis, **250,** 251, 385–386
Nucleus, **128**
NuSTAR space telescope, 157

O

O stars, 177
OB association, **230,** 319, 333
Oberon, A-8
Observable universe, **380**

Observatories
airborne and space, 116–118
optical and radio, 110–115
Observing the sky, A-11 to A-13
Octans (the Octant), 101
Olbers, Heinrich, 379
Olbers's paradox, **379**
1E 0657-56, 349
107 Piscium, 160
Opacity, **231**
Opebinia, 416
Open cluster, 253, 254, 313
Open orbit, 87
Open universe, 390, 393
Optical telescope, **107,** 110–115
Orbital motion, 84–89
Orbital velocity, 84–85
Orbits, 78, **84,** 86–87
Orion, 12, A-11
Orion Nebula, 221–225

P

Pagel, Bernard, 332
PAHs (polycyclic aromatic hydrocarbons), 206
Paradise Lost (Milton), 409
Parallax, 60, **171**
Parallax and distance, 172–173
Parsec (pc), **172,** A-5
Parsons, William, 337
Partial lunar eclipse, **38**
Partial solar eclipse, **40**
Particle astronomy, 123
Particle physics, 92
Paschen lines, 137
Paschen series, 137
Pasteur, Louis, **69**
Path of totality, **41**
Paul V, Pope, 72
Pauli exclusion principle, 247, 287
Payne, Cecilia, 149–150, 175
Peeled supernovae, 277
Pegasus, 13
Pelican, 326
Penumbra, **37,** 154
Penumbral lunar eclipse, **38**
Penzias, Arno, 383, 384
Perfection of the heavens, 56
Perigee, **42,** 87
Perihelion, 23
Period-luminosity relation, **253**
Periodic table of elements, A-10
Perlmutter, Saul, 397
Permitted orbits, **129**
Perseus A, 364
Perseus cluster, 364, 373
Philosophiae Naturalis Principia Mathematica (Newton), 91
Phobos, A-8
Phoebe, A-8
Photographic plate, **118**
Photometer, **118**
Photon, 103
Photon and particle soup, 385
Photosphere, **43, 144**–146
Pickering, James, 311
Pistol Star, 273
Planck space observatory, 117
Planck space telescope, 377, 402
Planck's constant, 103
Planetary nebulae, 265, 266–267
Planets, **3**
Plato, 55–56, 60
Pleiades star cluster, 203
Pliocene-Pleistocene mass extinction, 281
Pluto, A-8
Poe, Edgar Allan, 379
Polar axis, 113
Polaris, 14, 15, 20, 175, 178, 184, 253
Polarity, **153**
Polarization, 403
Police radar, 139
Polycyclic aromatic hydrocarbons (PAHs), 206
Poor clusters, **349,** 350
Pope, Alexander, 79
Population I stars, **328**
Population II stars, **328**
Positron, **163**
Post-main-sequence evolution, 246–251
Postulates of relativity, 93
Potential energy, 88
Powers of 10 notation, A-4 to A-5
Prajapati, 12
Precession, **20**–21
Pressure–temperature thermostat, **233**–234, 242
Primary lens, **106**
Primary mirror, **106,** 112
Prime focus, 112
Prime-focus cage, 112
Primordial soup, **414**
Principal satellites of Solar System, A-8
Principia (Newton), 91
Principle of cause and effect, 83
Principle of conservation of angular momentum, 288
Principle of relativity, 92, 93
Prism, 120
Process, 323
Procyon, A-6
Project Phoenix, 423
Projection, 21
Prominences, **43,** 158
Proper motion, **173**
Properties of planets, A-7
Proteins, 412, 413
Proteus, A-8
Protogalaxy, **330**
Protogalaxy gas cloud, 354
Proton, **128**
Proton–proton chain, **162**–164, 232
Protostar, **220,** 224
Protostellar disks, **226,** 228–229
Proxima Centauri, 263, 420, A-7
Prutenic Tables, 62, 64
Pseudoscience, 26
PSR 0540-69, 293
PSR B1257+12, 296, 297
PSR B1913+16, 293, 294
PSR B1957+20, 296
PSR J1748-2446ad, 295
Ptolemaic universe, 58
Ptolemy (Claudius Ptolemaeus), 14, 58, 61
Pulsar planets, 296–297
Pulsar wind, **292,** 293
Pulsars, 96
binary, 293–295
concept art, 290–291
defined, **289**
discovery of, 288–289
evolution, 289, 292
fastest, 295–296
lighthouse model, 289, 290–291
millisecond, 295
pulsar winds, 292, 293
Pulsating stars, 256–258
Pulsation period, 258
Puppis supernova remnant, 291
Pythagorus, 55

Q

Quantitative thinking, A-5
Quantum mechanics, **129,** 130
Quasar redshifts, 365, 368
Quasars, **365,** 368
Quintessence, **398**
QZ Vulpecula, 302

R

R Sculptoris, 264
Radial velocity (V_r), 139
Radiation, 102–104, 231
Radiation from heated object, 132–134

Radiation pressure, **230**
Radiative envelopes, 233, 234
Radiative zone, **164**
Radio galaxies, **362**
Radio intensity map, 120
Radio interferometer, 122
Radio lobes, 366–367
Radio map, 320, 326
Radio telescope, **107,** 115
Radio waves, 103, **104**
Rainbow, 103
Raisin bread analogy, 381
Re-ionization, **387**
Reasoning by analogy, 380
Recent supernovae (1987-2014), 280–281
Recombination, **387**
Reconnection event, 159
Red dwarfs, **181,** 182, 191, 192, 245, 263
Reddening curve, 204
Redshift, 138, **139,** 353, 365, 368, 389, 392
Reflecting telescope, **105**–107
Reflection nebula, 202, 203
Refracting telescope, **105**–106
Regular solids, 65
Regulus A, 183
Reinhold, Erasmus, 62
Reiss, Adam, 397
Relativity, 92–97
Renaissance, 74
Representational-color images, 119
Resolution and precision, 109
Resolving power, **108**
Retrograde motion, 60, 62
Revolution, **21**
Rhea, A-8
Rheticus, Joachim, 59
Ribonucleic acid (RNA), 413
Rich clusters, **349,** 350
Riemann, Bernhard, 393
Rigel, 15, A-6
  B8 star, 178
  color, 132, 134
  expanded and cooled, 246–247
  luminous supergiant, 184
  proper motion, 173
  winter sky, 14
Rigil Kentaurus, 15, A-6
Ring galaxy, 353
RNA (ribonucleic acid), 413
Roche lobe, **270**
Roche surface, **270**
Romeo & Juliet (Shakespeare), 33, 200
Ross 154, A-7
Ross 248, A-7
Rotation, **21**
Rotation curve, **318,** 345, 346
Rotation curve method, **346**
RR Lyrae variables, **253,** 256, 328
RU Camelopardalis (RU Cam), 258
Rubaiyat of Omar Khayyam, 128
Rudolphine Tables, 65, 66, 68–69
Russell, Henry Norris, 180
RX J185635-3754, 291, 293

S

S0 galaxies, 341, 356
Sa galaxies, 340
Sagan, Carl, 362, 424, A-1
Sagittarius, 12, 208, 312, 313
Sagittarius A* (Sgr A*), 323, 324, 326–327
Sagittarius Dwarf Galaxy (Sgr Dwarf), 350, 351
San Andreas fault, 186
Saros cycle, **47–**48
Saturn
  moons, A-8
  orbital properties, A-7
  physical properties, A-7
Sb galaxies, 340
SBa galaxies, 340
SBb galaxies, 340
Sbc galaxies, 340
SBc galaxies, 340
Sc galaxies, 340
Schmidt, Brian, 397
Schmidt, Maarten, 365
Schmidt-Cassegrain focus, 112
Schwarzschild, Karl, 298
Schwarzschild black hole, 298–299
Schwarzschild radius (R_S), **298**
Science, A-2
Scientific argument, **28**
Scientific confidence, 165
Scientific data, 194
Scientific imagination, 39
Scientific knowledge, 388
Scientific method, **7,** 8, 70
Scientific models, **17,** 241
Scientific notation, **3**
Scientific revolutions, 64
Scorpio, 12
Scorpius, 12, 313
Scott, David, 80
SDO (Solar Dynamics Observatory) spacecraft, 143, 157, 160
SDSS (Sloan Digital Sky Survey), 114–115
Search for extra-terrestrial intelligence (SETI), **422,** 423
Searching for natural causes, 268
Seasons, 21–24
Second postulate of special relativity, 93
Secondary mirror, 106, 112
Seeing, **108**–109
Segmented mirror, 113
Self-sustaining star formation, **322**–323
Semimajor axis *(a),* **67**
SETI (Search for extra-terrestrial intelligence), **422,** 423
SETI Institute, 423
SETI@home project, 423
Seyfert, Carl, 362
Seyfert galaxies, **362**–363
Sgr A*, 323, 324, 326–327
Sgr Dwarf (Sagittarius Dwarf Galaxy), 350, 351
Shakespeare, William, 24, 33, 171, 200
Shapley, Harlow, 312–315, 336, 337
Shapley-Curtis debate, **337**
Shock wave (shock), **212,** 218
SI *(Système International)* metric units, A-3
Sidereal drive, 113
Sidereal period, 35
Sidereus Nuncius (Galileo), 71, 72
Singularity, **298**
Sirius, 15, A-6
  A1 star, 177, 178
  main-sequence star (V), 184
  name, 13
  Temple of Isis, 53
  winter sky, 14
Sirius A, 187, 193, 265, A-7
Sirius B, 187, 193, 265, A-7
SKA (Square Kilometer Array), 122
Slipher, Vesto, **343**
Sloan Digital Sky Survey (SDSS), 114–115
Sloan Great Wall, 395
Slocum, Joshua, 12
Small-angle formula, **40**–41
Small galaxies, 356
Small Magellanic Cloud, 313, 341
SN 1006, 279
SN 1987A, 277, 280, 292
SN 2005cs, 276
SN 2005gl, 277
SN 2014J, 281
Snake, 326
Snow, John, 83
SNR 0509-67.5, 275
SO-2, 326
SOFIA (Stratospheric Observatory for Infrared Astronomy), 116, 281, 327
Solar and Heliospheric Observatory (SOHO), 148, 150
Solar constant, **160**
Solar Dynamics Observatory (SDO) spacecraft, 143, 157, 160

Solar eclipse, 38–45
annular eclipse, 42
apogee/perigee, 42
common misconception, 44
defined, **40**
diamond ring effect, 43
list of future eclipses, 43
observing an eclipse, 43–45
partial, 40
path of totality, 41
predicting eclipses, 45–48
total, 40
totality, 43
Solar flares, 159
Solar magnetic cycle, 152–156
Solar Max spacecraft, 148
Solar Maximum Mission satellite, 160
Solar neutrinos, 163, 164–166
Solar System, **3**
principal satellites, A-8
properties of planets, A-7
Solar TErrestrial RElations Observatory (STEREO) spacecraft, 160
Solar wind, **149**
Solstices, 22, 24
Sombrero, 335
South celestial pole, 18
South point, 18
Space telescope, 116–117
Special relativity, 92–94
Special theory of relativity, 298
Spectrograph, **119**–121
Spectral lines, **119**
Spectral sequence, **176,** 177
Spectral types, **176**–178
Spectral types—subtypes, 176
Spectroscopic binary system, **186**–187, 188
Spectroscopic parallax, **184**
Spectrum (spectra), 135–140
chemical composition, 135, 138
concept art, 136–137
defined, **103**
electromagnetic spectrum, 103–104
Kirchhoff's laws, 136
stellar spectra, 175–178
visible spectrum, 104
Spherical component, **317**
Spica, 14, 15
Spicules, **146,** 147
Spiral arms, **319**–323
Spiral density wave theory, **321**
Spiral galaxies, 339, **340,** 346
Spiral nebula, **337**
Spiral tracer, **319,** 321
Spitzer Space Telescope
Cas A, 279
deep-field images, 337
M101, 120
Milky Way Galaxy, 205, 324
Orion Nebula, 225
PSR B1257+12, 297
Ursa Major, 387
Spring tides, **90**
Spur, 320
Square Kilometer Array (SKA), 122
SRI International, 423
SS 433, 303
St. Augustine, 73
Standard candles, 342
Standard Model, 92
Star. *See* Stars
Star brightness, 14–16, 174, 193
Star charts, A-11, A-12 to A-13
Star clusters, 252–253, 254–255
Star distances, 171–173, 193
Star-formation pillars, **222**
Star gauging, 312
Star masses, 185–189, 191
Star names, 13–14
Star sizes, 179–184
Star system, 311
Star trails, 18
Starburst galaxies, **354**
Stardust spacecraft, 211
Stars, 170–198
absolute visual magnitude, 174–175
balance, 230–231
binary, 185–189
brightest, 193, A-6
brightness, 14–16, 174, 193
common misconceptions, 253
concept art, 192–193
convective cores/radiative envelopes, 233, 234
core, 233, 234
death. *See* Death of stars
defined, **3**
density, 194–195
diameter, 183
distances, 171–173, 193
dwarfs, 181
evolution. *See* Stellar evolution
famous, 14, 15
formation. *See* Formation and structure of stars
gas and dust, 212
giants, 181, 182
H–R diagram, 180–181
hydrostatic equilibrium, 230, 231
life expectancy, 245
luminosity, 175, 193
luminosity spectral classification, 183–184
magnitude and flux, 15–18
mass, 185–189, 191
mass-luminosity relation, 191, 194
names, 13–14
nearest, 193, A-7
proper motion, 173
radiative envelopes/convective cores, 233, 234
size, 179–184
spectral sequence, 176, 177
spectral types, 176–178
spectroscopic parallax, 184
spots and magnetic cycles, 160
stability, 230–231
structure, 230–232, 239, 240
supergiants, 181, 182
surveying, 191, 192–193
temperature spectral types, 176–178
trapezium, 221, 224, 225
Stars Are Yours, The (Pickering), 311
Starspots, 160
Statistical evidence, 364
Stefan, Josef, 133
Stefan-Boltzmann constant, 133
Stefan-Boltzmann law, 133, 134, 205
Stellar associations, 316
Stellar energy, 232–234
Stellar evolution, 238–261
concept art, 254–255
degenerate matter, 247–248
expansion into a giant, 246–247
fusing elements heavier than helium, 251
helium flash, 248, 249
helium fusion, 248–251
main sequence. *See* Main-sequence stars
mass loss, 264–265
nuclear reactions in massive stars, 252
nucleosynthesis, 250, 251
post-main-sequence evolution, 246–251
star clusters, 252–253, 254–255
triple alpha process, 248
variable stars. *See* Variable stars
Stellar life expectancies, 245
Stellar-mass black holes, 369
Stellar mass-luminosity relation, 191, 194
Stellar model, **240**–241
Stellar parallax *(p),* **172**
Stellar spectra, 175–178
Stellar structure, 230–232, 239, 240
Stellar winds, **230**
STEREO (Solar TErrestrial RElations Observatory) spacecraft, 160
Stonehenge, 53, 54
Stratospheric Observatory for Infrared Astronomy (SOFIA), 116, 281, 327
Stromatolites, **411**

Strong nuclear force, **162,** 163
Subclustering, 350
Subdwarfs, 184
Subgiants, 184
Sudbury Neutrino Observatory, 165
Summer solstice, 22
Sun, 143–169, A-6, A-7
  abundances of elements, 150
  annual motion, 21
  chromosphere, 146–147
  chromospheric and coronal activity, 156–160
  composition, 149–150
  concept art, 153–154, 158–159
  convective zone, 164
  corona, 147–149
  energy transport, 164
  G2 star, 177
  helioseismology, 150, 151
  hydrogen fusion, 162–164
  magnetic cycle, 152–156
  meridional flow, 156
  neutrinos, 163, 164–166
  nuclear fusion, 161–165
  observing, 151
  photosphere, 144–146
  radiative zone, 164
  solar constant, 160
  solar eclipse, 38–45
  solar wind, 149
  spectrum, 135
  stellar model, 240
  sunspots, 152, 154–155
Sun Dagger, 53, 54
Sunspots, **151,** 152, 154–155
Suomi National Polar-orbiting Partnership satellite, 1
Superclusters, **395**
Superflares, 160
Supergiant, **181,** 182, 195, 247
Supergranules, **145**
Supermassive black holes, 369–374
  disks and jets, 369
  history of, 372–374
  mass, 346–347
  Sgr A*, 324
  triggering eruptions, 371
  unified model of active galaxies, 369–370
Supernova, **274**
Supernova explosions, 274–281
  carbon deflagration, 277
  Crab Nebula, 278
  Earth, and, 281
  exploding core of supernova, 276
  historical supernovae, 278–280
  iron core, 274
  Kepler's supernova of 1604, 278, 280
  models of supernovae, 274–275
  observations of supernovae, 274
  recent supernovae (1987-2014), 280–281
  supernova remnants, 274, 279
  type I supernova, 275
  type Ia supernova, 277
  type Ib supernova, 277
  type II supernova, 275, 277
Supernova remnants, **274,** 279
Superwind, **264**
Surveyor's triangulation method, 171
Swift orbiting telescope, 305
Synchrotron radiation, 278
Synodic period, 35

T

T association, **230**
T dwarfs, 178, 179, 243
T Leporis, 123
T Tauri stars, 228
Tarantula Nebula, 341
Taurus the Bull, 278
Taylor, Joseph, 293, 294
Telescope
  airborne, 116
  diffraction fringes, 108
  light-gathering power, 107–108
  magnifying power, 110
  optical, 110–115
  radio, 115
  refracting/reflecting, 105–107
  resolving power, 108–109
  seeing, 108–109
  space, 116–117
Telescope resolution, 108
Temperature, **132**
Temperature scales, 132, A-4
Temperature spectral types, 176–178
Temple of Isis, 53
Termination shock, 211
Terrible rule of evidence, 388
Tethys, A-8
Thales of Miletus, 47
Theorist, 70
Theory, **69,** 221
Thermal energy, **132**
Thermophile, **418**
Theta Leonis, 15
Third postulate of relativity, 93
30 Doradus, 222
Thirty Meter Telescope (TMT), 114
Thomson, J. J., 39
Thomson, William (Lord Kelvin), A-4
Thor's Helmet, 242
Threads, 326
3C 31, 363–365, 367
3C 58, 291
3C 75, 367
3C 273, 368
Through Rugged Ways to the Stars (Shapley), 336
Thymine (T), 412
Tianyan, "The Eye of Heaven," 115
Tidal distortion, 352
Tidal tails, 352, 363, 372
Tides and tidal forces, 89–91
Time, 6–7
Time dilation, **299**
Titan, 418, A-8
Titania, A-8
TMT (Thirty Meter Telescope), 114
Top-down hypothesis, **330**–331
Total lunar eclipse, **37**–38
Total solar eclipse, **40**
Totality, **37**
TRACE spacecraft, 158
Transition, 137
Trapezium stars, 221, 224, 225
Triangulation method, 171–172
Trifid Nebula, 203
Trilobites, 416
Triple alpha process, **248**
Triton, A-8
Turnoff point, 254
Twain, Mark, 217
21-cm radiation, **206**–207
21-cm radio map, 320
Twin Keck telescopes, 113, 114, 123
Two-Degree-Field Redshift Survey, 397
Two New Sciences (Galileo), 81
Tycho's supernova, 64, 279
Tycho's supernova remnant, 120
Type 1 Seyfert galaxies, 362
Type 2 Seyfert galaxies, 362
Type I Cepheids, 253, 256, 258
Type I supernova, 275
Type Ia supernova, 277, 342, 343, 399
Type Ib supernova, 277
Type II Cepheids, 256, 258, 328
Type II supernova, 275, 277

U

UFOs (unidentified flying objects), 420, 421
Ultraluminous infrared galaxy (ULIRG), **355**
Ultraviolet (UV) radiation, **104**
Umbra, **36,** 37, 154
Umbriel, A-8

Unidentified flying objects (UFOs), 420, 421
Unified model, **370**
Uniform circular motion, **56,** 60
Units and astronomical data
  astronomical constants, A-5
  brightest stars, A-6
  conversions, A-4
  Greek alphabet, A-9
  main-sequence stars, A-6
  meteor showers, A-9
  metric prefixes, A-3
  nearest stars, A-7
  periodic table of elements, A-10
  powers of 10 notation, A-4 to A-5
  principal satellites of Solar System, A-8
  properties of planets, A-7
  SI units, A-3
  temperature scales, A-4
  units used in astronomy, A-5
Universal mutual gravitation, 83
Universe. *See* Cosmology
Upper-main-sequence O stars, 191
Upper-main-sequence stars, 263. *See also* High-mass stars
Uranus
  moons, A-8
  orbital properties, A-7
  physical properties, A-7
Urban VII, Pope, 72
Urey, Harold, 411
Ursa Major, 13
UV (ultraviolet) radiation, **104**

V

V404 Cygni, 302
Vacuum of space, 200
van Arkel, Hanny, 372
Van der Waals radius, 128
Variable stars, 253–258
  Cepheids, 253, 256–258
  defined, **253**
  intrinsic variables, 253
  period changes, 258
  period-luminosity relation, 253
  pulsating stars, 256–258
  RR Lyrae variables, 253, 256
  type I Cepheids, 253, 256, 258
  type II Cepheids, 256, 258
Vega, 15, A-6
  A0 star, 178, 183
  main-sequence star (V), 184
  proper motion, 173
  size, 183
  visible on summer and fall evenings, 14
Vela pulsar, 292, 293
Vela supernova remnant, 293
Velocity, **81,** 138–139
Velocity dispersion method, **346**
Venus, 25, A-7
Vernal equinox, 22
Very Large Array (VLA), 122
Very Large Telescope (VLT), 111, 114, 122
Very Long Baseline Array (VLBA), 122
Viking spacecraft, 419
Violent motions, **79,** 80
Virgo cluster, 349, 380, 395
Virus, 323
Visible spectrum, 104
Visual binary system, **186**
VLA (Very Large Array), 122
VLBA (Very Long Baseline Array), 122
VLT (Very Large Telescope), 111, 114, 122
VLTI, 122–123
Voids, **395**
von Fraunhofer, Joseph, 128
Voyager spacecraft
  heliopause, 149
  interstellar medium (ISM), 211
Voyager 1 spacecraft, A-1
VW Cephei, 190
VY Canis Majoris, 269

W

Waning crescent moon, 35
Waning gibbous moon, 35
Water hole, **422**
Wavelength, **102**
Wavelength–frequency–speed equation, 103
Wavelength of maximum intensity (λ_{max}), 134
Waxing crescent moon, 34
Waxing gibbous moon, 34
Weak nuclear force, **162,** 163
Weather map, 120
Wekiu bugs, 418
West point, 18
Westerlund 1, 223
WFIRST (Wide-Field Infrared Survey Telescope), 117
Whirlpool Galaxy, 352
White dwarf, **181,** 182, 195, 265–269
White dwarf remnants, 263
White Tiger, 12
Wide-Field Infrared Survey Telescope (WFIRST), 117
Wien, Wilhelm, 134
Wien's law, 133, 134
Wilkinson Microwave Anisotropy Probe (WMAP), 402
Wilson, Robert, 383, 384
WIMPs, **394**
Winter solstice, 22
WMAP (Wilkinson Microwave Anisotropy Probe), 402
Wolf 359, A-7
Wolf-Rayet star, 242
World Wide Telescope, 115
WR 124, 269
Wright, Thomas, 312

X

X Cygni, 258
X-ray, **104**
X-ray bursters, **295**
X-ray telescope, 117–118
XMM-Newton, 117
XTE J1751-305, 295

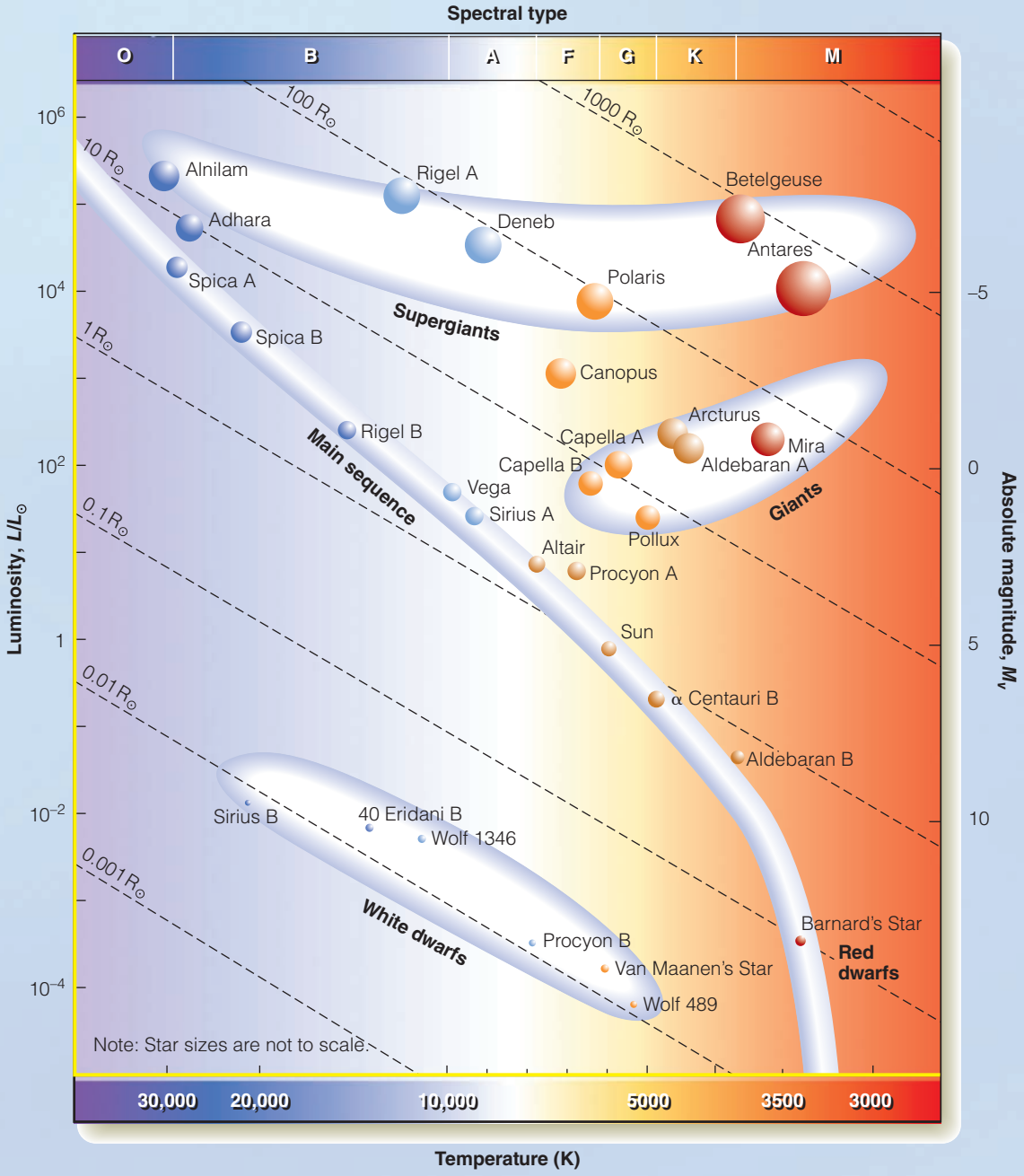
Y

Y dwarfs, 178
Yellow jacket, 157
Yellowstone National Park, 418
Young stellar associations, 230
Young stellar objects (YSOs), **226,** 228–229

Z

Zeeman effect, 155
Zenith, 18
Zero-age main sequence (ZAMS), **244**
Zodiac, **25**
ZwCl 0024.0+1654, 349
Zwicky, Fritz, 287

Flash Reference:
H-R Diagram



The H–R diagram is the key to understanding stars, their birth, their long lives, and their eventual deaths. Luminosity ($L / L_{\odot}$) refers to the total amount of energy that a star emits in terms of the Sun's luminosity, and the temperature refers to the temperature of its surface. Together, the temperature and luminosity of a star locate it on the H–R diagram and tell astronomers its radius, its family relationships with other stars, and a great deal about its history and fate.

Flash Reference: Comparative Planetology

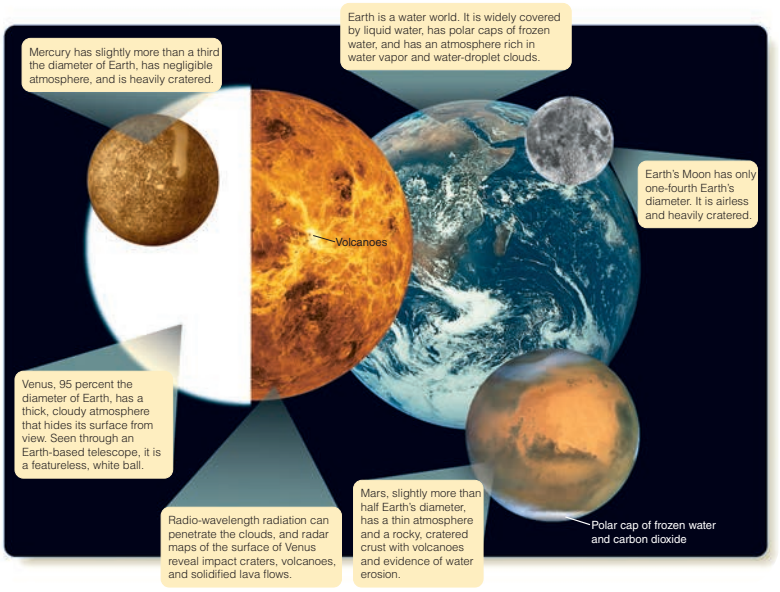
The Terrestrial or Earthlike planets orbit very close to the Sun, and their orbits are hardly visible in a diagram that includes the outer planets.

Mercury, Venus, Earth and its Moon, and Mars are small worlds made of rock and metal with little or no atmospheric gases.

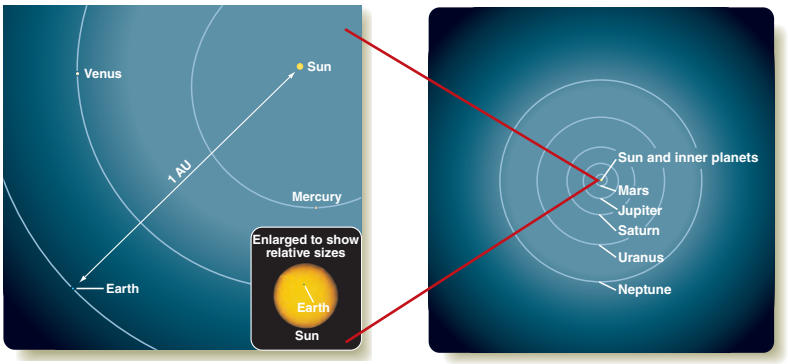
The outer worlds of our Solar System orbit far from the Sun. Jupiter, Saturn, Uranus, and Neptune are Jovian or Jupiter-like planets much bigger than Earth. They contain large amounts of low-density gases.

Pluto is one of a number of small, icy worlds orbiting beyond Neptune. Astronomers have concluded that Pluto is more appropriately labeled a dwarf planet.

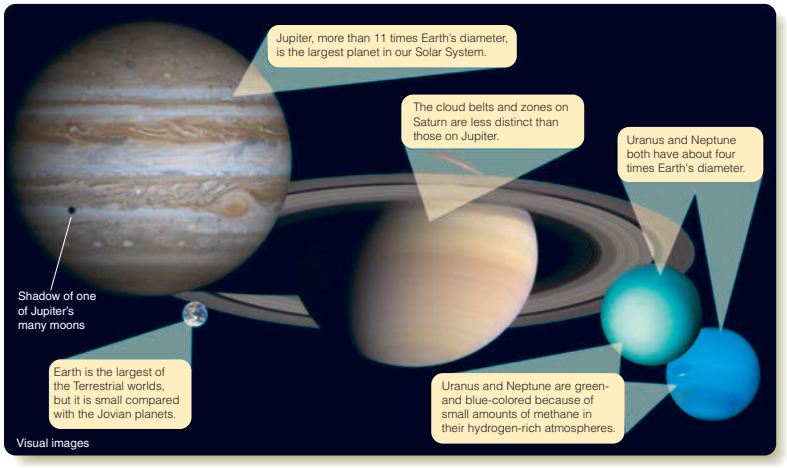
The Terrestrial Worlds


Mercury has slightly more than a third the diameter of Earth, has negligible atmosphere, and is heavily cratered.
Earth is a water world. It is widely covered by liquid water, has polar caps of frozen water, and has an atmosphere rich in water vapor and water-droplet clouds.
Earth's Moon has only one-fourth Earth's diameter. It is airless and heavily cratered.
Volcanoes
Venus, 95 percent the diameter of Earth, has a thick, cloudy atmosphere that hides its surface from view. Seen through an Earth-based telescope, it is a featureless, white ball.
Radio-wavelength radiation can penetrate the clouds, and radar maps of the surface of Venus reveal impact craters, volcanoes, and solidified lava flows.
Mars, slightly more than half Earth's diameter, has a thin atmosphere and a rocky, cratered crust with volcanoes and evidence of water erosion.
Polar cap of frozen water and carbon dioxide

Planetary Orbits



The Outer Worlds


Jupiter, more than 11 times Earth's diameter, is the largest planet in our Solar System.
The cloud belts and zones on Saturn are less distinct than those on Jupiter.
Uranus and Neptune both have about four times Earth's diameter.
Shadow of one of Jupiter's many moons
Earth is the largest of the Terrestrial worlds, but it is small compared with the Jovian planets.
Uranus and Neptune are green- and blue-colored because of small amounts of methane in their hydrogen-rich atmospheres.
Visual images

Flash Reference: Arrows

This book is designed to use arrows to alert you to important concepts in diagrams and graphs. Some arrows point things out, but others represent motion, force, or even the flow of light. Look at arrows in the book carefully and use this Flash Reference card to catch all of the arrow clues.

Point at things:

You are here

Force:

Process flow:

Measurement:

Direction:

Radio waves, infrared, photons:

Motion:

Light flow:

Focal length